Atlantic Council DFRLab

PROTECTING POINT-TO-POINT MESSAGING APPS

Understanding Telegram, WeChat, and WhatsApp in the United States

By Iria Puyosa

The mission of the **Digital Forensic Research Lab (DFRLab)** is to identify, expose, and explain disinformation where and when it occurs using open-source research; to promote objective truth as a foundation of government for and by people; to protect democratic institutions and norms from those who would seek to undermine them in the digital engagement space; to create a new model of expertise adapted for impact and real-world results; and to forge digital resilience at a time when humans are more interconnected than at any point in history, by building the world's leading hub of digital forensic analysts tracking events in governance, technology, and security.

Author

Iria Puyosa

Editors

Graham Brookie
Andy Carvin
Rose Jackson
Iain Robertson

Acknowledgements (Sponsorship):

This report was made possible with support from the Omidyar Network.

Atlantic Council
1030 15th Street NW, 12th Floor
Washington, DC 20005

For more information, please visit
www.AtlanticCouncil.org.

July 2023

Design by: Donald Partyka and Anais Gonzalez

Cover: Created by Anais Gonzalez.

Atlantic Council DFRLab

PROTECTING POINT-TO-POINT MESSAGING APPS

Understanding Telegram, WeChat, and WhatsApp in the United States

By Iria Puyosa

Table Of Contents

Executive Summary

Too often, consideration of point-to-point messaging platforms in the United States is focused on either diaspora or second-language usage, given the global popularity of these platforms. Another common focus is on extremist or unlawful usage. In reality, a broad swath of Americans use point-to-point platforms, the popularity of which is increasing, but that usage remains at a lower rate when compared to that in other regions of the world. An estimated 69 percent of the United States population currently uses at least one point-to-point messaging app, though the use and dynamics of this part of the information ecosystem remain understudied.

The Digital Forensic Research Lab (DFRLab) undertook this project to better understand and contextualize point-to-point platform usage in the United States with two goals: first, to analyze the growing use of these platforms in the United States; and, second, to emphasize the growing importance of rights respecting—and protecting—elements of some platforms, such as end-to-end encryption as an important technology at the core of designing for data privacy and free speech.

The DFRLab carried out this research project to shed light on the following topics:

- First, how point-to-point platforms work, their varying degrees of security features, and how they deploy encryption.
- Second, understanding how diverse communities use the messaging platforms for different purposes.
- Third, understanding the variance among platform design and enforcement of terms of usage.
- Finally, how messaging app security is important for protecting and respecting rights—like privacy and freedom of expression—in this digital era.

We mapped the ecosystem of point-to-point messaging apps in the United States, looking at the more than forty apps available in the market. We assessed the features these apps offer, their registration requirements, and their approach toward encryption.[1]

The messaging apps reviewed may be similar in communication features but varied substantially in security, privacy, and content policies. The intersection of technical features, policies, and detection methods around acceptable usage (as defined by the platforms) leads to different models for use. Ultimately, we chose to focus our empirical research on Telegram, WeChat, and WhatsApp because they present distinct product architectures and technical features, and varying policies on usage.

Platforms must balance complex trade-offs to protect their users and ensure app integrity. Messaging apps typically establish policies of acceptable usage, prohibiting some harmful or criminal content, ranging from spam to sexual abuse material and terrorism. Telegram has a permissive content policy, but the platform has been adding restrictions in recent years following pressure from law enforcement in different countries. WhatsApp has a growing list of unacceptable content considered harmful or illegal. WeChat is the most restrictive messaging app regarding acceptable content, banning even political content. All three of these messaging apps prohibit sharing content depicting sexual abuse or calls for violent crimes.

Messaging app security depends on how encryption is enabled. Almost every messaging app offers data encryption in transit between devices, as is standard in most internet-enabled data exchanges. Additionally, most reliable messaging apps provide end-to-end (E2E) encryption, which protects messages from unauthorized access by third parties, including the platform itself. WhatsApp offers E2E encryption by default, Telegram offers opt-in encryption, and WeChat only offers transport-layer encryption for data in transit.

In general, data collection is less extensive in messaging apps than on mainstream social media platforms such as Twitter or Facebook. Few messaging apps conduct extensive monitoring for unacceptable content since human moderation and automated scanning would infringe on their terms of service. However, most messaging apps collect basic usage metadata to monitor platform performance and integrity. Telegram collects minimal usage data, WhatsApp collects sizable usage data, and WeChat extensively captures both usage and content data. As

1 A comparative table of the fifteen apps most commonly used for direct messaging in the United States is presented on page 11.

such, Telegram and WeChat are, in many ways, at opposite ends of the spectrum, where Telegram is loosely moderated and controlled while WeChat comprehensively tracks its users, their behavior, and the content they post.

Remarkable differences exist among the three messaging platforms that the DFRLab focused on in this report. Telegram's design prioritizes that the content of communications be available on different devices. Its public channels offer large group sizes, ample reach, and many features for reacting to content. WeChat is an all-encompassing app in which interaction with service and official accounts is paramount. Automated monitoring to ensure compliance with its policies of acceptable content is built into the design, in compliance with Chinese regulations. WhatsApp's original design aimed to satisfy the needs of direct individual-to-individual personal communications. Thus, it still favors a balance between privacy and safety, although this may change as the platform embraces other forms of interactions, such as communities, public channels, and business transactions.

Usage of messaging platforms is growing and overwhelmingly lawful and beneficial. The DFRLab observed the following general trends:

- Messaging conversations often link to content posted on social media platforms and the open web.
- Local communities' dynamics and information related to transnational issues are intertwined.
- Diaspora communities rely on WhatsApp and WeChat for mutual support and exchange of resources.

The case studies in this report were selected as illustrations of a cross section of platforms and communities or uses that have either received extensive news coverage or too little. In our analysis, we found different ways in which misinformation and foreign influence operations spread—or did not spread—on Telegram, WeChat, and WhatsApp. We found that political or ideological topics were more prevalent in messaging interactions among US-born users in public Telegram groups than among foreign-born diaspora communities. Moreover, we observed issues outside our initial scope. These issues included intrusive practices such as business spamming and outright harms such as the unsolicited posting of sexual abuse content on public groups. Upon analysis of public groups and channels on WhatsApp, Telegram, and WeChat, the DFRLab observed the following outlying findings:

- Misinformation and disinformation about political and health topics were widespread on the public Telegram channels, health-related misinformation was found in WhatsApp public groups, and misleading political narratives were detected on WeChat public accounts.
- Individuals and groups in the United States who espouse white supremacist beliefs are active on Telegram public channels in a way that they are not able to be (under the terms and conditions) on larger social media platforms like Facebook or Twitter.
- Public WeChat accounts were instrumentalized to foster narratives aligned with the Chinese Communist Party among various groups.
- Pro-Russian influence campaigns were active on public Telegram channels in English and Spanish.
- Supporters of former US President Donald Trump used public Telegram channels to boost their political views ahead of the 2022 midterm elections, and they are already sharing content related to the 2024 presidential elections.
- Unsolicited sharing of sexual imagery and content derived from sexual exploitation, including child sexual abuse, was found in a few public WhatsApp groups.
- Some users with business accounts violate WhatsApp's acceptable usage policies by engaging in spam, offering prohibited transactions such as cryptocurrencies, or advertising fraudulent products.

Messaging platforms can rely on methods that do not require accessing message texts or images in compliance with policies and terms of usage. These methods are in-app user reporting, analysis of metadata, and analysis of behavioral signals. WhatsApp uses all three methods for enforcing its policies. Telegram relies mainly on in-app user reporting, although the platform has capabilities for metadata analysis. WeChat also encourages user reporting, but this platform deploys automated content scanning for interactions within the app.

Some organizations working on counterterrorism or child sexual abuse have been asking for privileged access or backdoors for law enforcement and deployment of automated scanning in messaging apps. E2E encryption renders automated scanning of content impossible, making it equally impossible for E2E encrypted apps to implement many common content policies of more open platforms, since they cannot decrypt content shared by

their users. Content-dependent preemptive methods, such as server-side or client-side scanning to match content a user is sending against a database, compromise encryption integrity, weaken security, and erode privacy protection. Both server-side and client-side scanning are ineffective for identifying never-seen-before content that is not already part of a database. Currently, "hashes" databases are available for terrorist content and child sexual abuse material posted on social media.

Security experts warn that automated content scanning undermines encryption and introduces security vulnerabilities in messaging apps, increasing risks for all users. Conversely, machine-learning procedures applied to metadata and behavioral signals would not compromise encryption and may detect never-before-seen content.

Based upon this investigation, the DFRLab recommends that platforms prioritize the following:

- Investing in in-app reporting tools.
- Defining robust policies for business and organizational accounts.
- Partnering with outside researchers to investigate the spread of harmful content, while establishing protocols for protecting users' personal data in the process.
- Collaborating with counterterrorism hashes databases.
- Considering impacts on human rights when designing policies and products.

Likewise, the DFRLab recommends that policymakers prioritize the following:

- Enacting data privacy protection legislation.
- Avoiding regulations that undermine rights-protecting technologies, such as E2E encryption.
- Examining business practices and commercial services offered via messaging apps to identify regulatory gaps.
- Promoting digital literacy tailored to the risks faced by users of messaging apps.

As an underlying ethos, legislators and policymakers should always take into consideration how policies and regulations aiming to govern or control messaging apps could be enforced across countries that maintain different levels of respect for human rights. For instance, a regulation instituted in the United States that mandates platforms keep identification records for their users and deliver that information to law enforcement agencies upon request could be weaponized in authoritarian or autocratic countries where a given messaging app is widely used, increasing the possibility of capture and incarceration of political dissidents. Similarly, requiring messaging apps to build in means for privileged access to E2E encrypted communications in a domestic context would likely open the door for other governments to repurpose the same technical infrastructure for surveillance. Ultimately, all actions taken by any company or government have potential impact beyond their intended target, often creating unintentional harm, and this potential must be a persistent consideration in every decision about how an app should operate.

Introduction

Point-to-point messaging apps[2] underpin the daily interactions of more than three billion people worldwide. They have exploded in popularity in Latin America, South Asia, Africa, and most European countries, and have a growing user base in the United States. Messaging apps are deeply embedded in daily life as a primary means of communication with friends and family, buying and selling products and services, following the news, and discussing public affairs.

The United States has lagged in adopting messaging apps due to the prevalence of affordable text messaging via mobile networks and greater reliance on direct messaging via mainstream social media platforms. Nonetheless, the usage of point-to-point messaging apps in the United States has grown significantly in recent years. A great deal of the initial growth was driven by diaspora communities using these apps to keep in touch with their friends and relatives abroad, who already relied on these tools. Recently, more US-born people have begun to use closed messaging apps. As of September 2021, 81.5 percent of the US adult population using mobile phones reported having a messaging app installed.[3]

The increasing user base has also provoked a surge of public debate about messaging apps centered on potential harms such as the spread of misinformation and disinformation, as well as other illegal activities previously occurring offline or through social media. Calls for content moderation within messaging apps have been fueled by news coverage on incidents of harmful usage. Meanwhile, law enforcement agencies have demanded privileged access, alleging national security concerns or supposed criminal activity hidden in private conversations. The outcry has often been compounded by a lack of understanding of how people use messaging and how encryption works to protect personal data and prevent unauthorized access to private communications.

The United States has a lot of room to grow regarding the extent to which messaging apps are used, compared with other countries. Still, a shift is underway, and we expect increasing adoption in the near future. To prepare for significant user base growth and exploding messaging app popularity in the United States, the Atlantic Council's Digital Forensic Research Lab (DFRLab) initiated a research project to explore how these apps are effectively being used in the United States, through the examination of a subset of apps and demographics, particularly diaspora communities. The resulting report aims to shed light on how individuals adjust their usage to features and policies and to discern content, privacy, and safety policies enacted by the three investigated apps. This work was done with care for ethical, transparent, and replicable methodologies, as well as an emphasis on features within platforms that protect fundamental rights like privacy and free speech in a digital age.

2 Although there are minor differences in the terms "closed messaging" and "point-to-point messaging," we will use the terms interchangeably in this report.

3 Statist, "Mobile Internet Usage in the United States," Statista Global Consumer Survey (GCS), 2022.

Methodology

Initially, our investigation aimed to understand the spread of misinformation and foreign influence operations on messaging apps, especially among diaspora communities. To do so, we first mapped the ecosystem of point-to-point messaging applications in the United States to identify the most commonly used messaging apps, the product features they offer, their registration requirements, and their approach toward encryption. The results of this mapping comprise the first subsection below, Messaging Apps Overview: Product Features, Privacy, and Security.

After the mapping, the DFRLab chose to focus on Telegram, WeChat, and WhatsApp, because of their differing usage in terms of demographics and their differentiated security models and acceptable content policies. Additionally, these three messaging apps offer *public* groups commonly used for sharing news and discussing public affairs in an open forum, which allows researchers to transparently and ethically observe usage dynamics. Conversely, researching *private* groups without the consent of users within the group raises significant ethical concerns.[4]

Given that we sought to observe how diaspora communities and other demographic groups take advantage of these apps for connecting language, identity, and affinity, the DFRLab intentionally limited the scope to public groups and channels on the three messaging apps, which we analyzed to shed light on how product features, security and privacy, acceptable use policies, and user behavior come into play in these closed messaging ecosystems.

The DFRLab selected three case studies to explore the use of messaging apps by diverse demographic groups, including those debating current events in the United States on Telegram, Chinese students in US higher education on WeChat, and Latino populations on WhatsApp. While identifying public groups on WhatsApp for research, we added a set of groups targeting a more general English-speaking US population. This mix of messaging apps and user demographics allowed us to analyze issues arising within different communication ecosystems. The DFRLab conducted its research from December 2021 through June 2022.

In parallel to these efforts, the DFRLab assessed platforms' terms of service, privacy policies, and acceptable content or usage policies. The assessment of the policies helps us to put in context the patterns of content-sharing observed during our research. Upon the evidence of harmful content circulating on messaging apps, we analyzed the current methods employed by platforms to monitor and enforce their policies. This allowed us to analyze the methods and practices available for enforcing policies on acceptable content in relation with observed usage and detected violations. At the end of the report, we also discuss methods being pushed forward for groups with special interests.

The data gathering and analysis methodology varied among the three case studies due to different challenges for assessing groups and building structured datasets. The report does not aim to provide statistical comparisons across platforms or generalize beyond the observed samples.

4 Connie Moon Sehat and Tarunima Prabhakar, "Ethical Approaches to Closed Messaging Research: Considerations in Democratic Contexts," *MisinfoCon Elections Conference Proceedings,* 2021.

Messaging Apps Overview: Product Features, Privacy, and Security

The DFRLab embarked on this project to assess the messaging app environment in the United States. Our research started by identifying the messaging apps commonly used and assessed their primary conversational, privacy, and security features as a means of identifying which apps merited further study.

While the encryption of any given messaging app tends to be the focus of conversation around these platforms, their users are more likely to be drawn to the app because of their prevalence within their demographic community; for example, WhatsApp's popularity in much of Latin America drives its popularity within the Latino diaspora in the United States as well. That said, human rights activists under threat, for example, are more likely to consider the security of any given app as a primary factor in their decision whether to use it.

When it comes to uptake in the United States, nearly forty messaging apps compete for market share. Messaging usage is not exclusive, however, and many people use several apps for complementary functions. The most popular messaging apps are WhatsApp, Telegram, and iMessage. Other popular direct messaging apps are embedded in social media platforms, such as Facebook Messenger, WeChat, Snapchat, and gaming platform Discord's instant messaging. Some other apps are less popular but are considered more secure due to their encryption protocols, such as Signal, Wire, Wickr, and Threema.

Product Features

Closed messaging apps offer synchronous and asynchronous conversations. Conversations can be one-to-one or within small or large groups. The modes of interactions within a messaging app can be two individuals; an individual with one organization; small- or medium-size closed groups in which most participants know each other; or large public groups to which anyone can join by obtaining the corresponding link. Small private groups are allowed in most messaging apps, while large public groups are allowed in WhatsApp, WeChat, and Telegram (up to 1,024 members on WhatsApp and 200,000 on Telegram).

Messaging apps' conversational features do not differ much, as developers mimic innovations from each other. Almost all messaging apps allow for sending multimedia, including audio, video, pictures, emojis, and stickers. Most messaging apps allow file attachments and clickable links. These features enable messaging chats to become a stream of news, resources, and commentaries, as individuals chatting share links to social media posts, news stories, and websites, commenting on them, and pointing out agreements and disagreements. Chat participants can attach files from their phones or computers, including large documents and even books in .pdf format, to share information and back their points. Thus, an active group chat can become a gateway to a wide array of media resources and social media conversations. Some messaging apps also enable easy sharing or forwarding of content to other users within the same app and even to other apps, as is the case with WhatsApp and Telegram.

Although most of the exchanges in messaging apps continue to be one-to-one generally, other conversations are increasingly taking place on messaging apps, including public groups, news distribution, and business-to-customer interactions. Platforms that offer differentiating features for individual users and organizational or business accounts regularly enact policies that are specific to these sorts of customers. In most cases, business or organizational accounts are expected to present themselves clearly as such.

Data Privacy

Point-to-point or closed messaging apps offer several privacy protection features. The highest level of data privacy is provided by apps that do not collect user data at registration, restrict metadata collection to the minimum required to deliver messages, and delete data from their servers after messages are delivered. The premise is that, if metadata is never collected, it can never be misused to violate user privacy and put them at risk. Nonetheless, most messaging apps require some form of user identification and collect basic usage metadata for monitoring platform performance and integrity. Other mainstream messaging apps require real user identification and may collect extensive usage metadata. The less

privacy-oriented platforms require real user identification and collect extensive usage metadata; some platforms store data in their servers and conduct metadata analysis for different purposes.

User identification—ranging from just a user-generated account name to a government issued ID—is the first step toward privacy in messaging apps. Most messaging apps require a phone number for registration and continuous usage. However, for some apps, users can register with a "burner" number[5] or with a disposable email address, neither of which are necessary to continue using after a username or user ID has been created. The most privacy-oriented registration requirements are those of Wickr and Threema, which only require downloading the app and generating a user ID. Messaging apps that manage user identification through user-generated IDs or username may provide anonymity to their users. That is the case with Wire, Wickr, Threema, and Telegram. Nonetheless, there are messaging apps that manage user identity using IDs, but identity is still linked to the phone number given for registration, as happens with WeChat. Less privacy-oriented registration requirements are those of Facebook Messenger, which requires a personal Facebook account. Therefore, it would link to the user's complete social graph in the parent app.[6]

How platforms handle registration and user identification have consequences on privacy risks. For instance, since a Telegram user's identity is a self-generated username, one can potentially only see anonymous usernames in a public Telegram channel. In contrast, participants in a WhatsApp or a Signal group can see the phone number of every other member. Since phone numbers can be tracked to the person subscribing to that phone line, there is a risk of revealing user identity on Signal and WhatsApp groups. Hence, digital rights activists have been pressuring Signal to switch to username or user key. In response, Signal President Meredith Whittaker declared in December 2022 that the platform is "hopeful that [usernames will] launch in the first half of 2023."[7]

Most point-to-point or closed messaging apps offer additional privacy protection features, such as blocking undesired contacts and ephemeral (or disappearing) messages. Most messaging apps also offer the ability to block spam and undesirable contacts. Wickr, Kik, and Snapchat offer ephemeral messages by default. In most other messaging apps, the sender has to opt into disappearing or ephemeral messages. The duration of the message before disappearing may vary depending on the setting selected, as it is on Signal, WhatsApp, Wire, and Telegram. Ephemeral messages disappear from all end devices and from servers after the set time.

Security

The most crucial part of the security configuration in a messaging app is whether and how encryption is enabled. Messaging apps offer different configurations for transport layer security (TLS) and end-to-end (E2E) encryption.

TLS is almost omnipresent in the lives of those using the internet, whether they know it or not. In short, TLS is a cryptographic protocol that encrypts data sent between applications over the internet.[8] Everyday activities protected by TLS encompass account logins, online banking, online shopping transactions, credit card payments, and browsing secure websites (those starting with *https*). The protocol, however, does not secure data in endpoints or servers, only during transit. Those who have access to the server, such as the platform owner, have access to communications content, enabling them to potentially share that content with third parties such as law enforcement.

Unencrypted servers are more vulnerable to hacking and data breaches, as, without encryption, any malicious actor who obtains unauthorized access to a server can read and potentially exfiltrate any data stored within it. Unauthorized access to content stored in app servers can

5 A "burner" phone number is a number that a person can lease temporarily in order to receive calls or messages without giving one's direct phone number. One use for burner phone numbers, as it pertains to this report, is registering with apps and web services that only require this information at the moment of setting up a new account.

6 A "social graph" on Facebook comprises all the friends' connections [and friends of friends] that the user has on Facebook, plus all the fan pages they like or the groups they join on Facebook.

7 Billy Perrigo, "Signal's President Meredith Whittaker Shares What's Next for the Private Messaging App," *Time*, December 4, 2022, https://time.com/6238482/signals-president-meredith-whittaker-interview/.

8 For a more detailed explanation of TLS and how it works, see "TLS Basics," Internet Society, n.d., https://www.internetsociety.org/deploy360/tls/basics/.

also be mitigated by encryption at rest (i.e., encryption of content stored on the servers), as offered on iMessage and in WhatsApp's cloud backups. In basic terms, encryption converts messages from plain text into ciphered data, rendering messages unreadable while stored and thus protecting the information from access by the platform, governments, or hackers.

E2E encryption ciphers messages so that only the sender and the intended receiver can decipher them, preventing even those with access to the app server from reading the message. E2E encryption protects user communication from access by an unauthorized party, including the platform itself or malicious actors such as terrorist organizations or repressive governments. Platforms may opt to provide E2E encryption by default, as it offers the highest level of such security. Some apps offer opt-in E2E encryption, giving users the choice between the ability to access their communications on multiple devices and the ability to secure communications from third-party access. Other platforms do not provide encryption in servers or endpoints, only in-transit encryption using TLS. Some platforms offer opt-in encrypted backups, with the conditions for backups varying. Meanwhile, other platforms do not securely encrypt servers, and some platforms indicate that they do not keep any content in their servers after delivering to the end point.

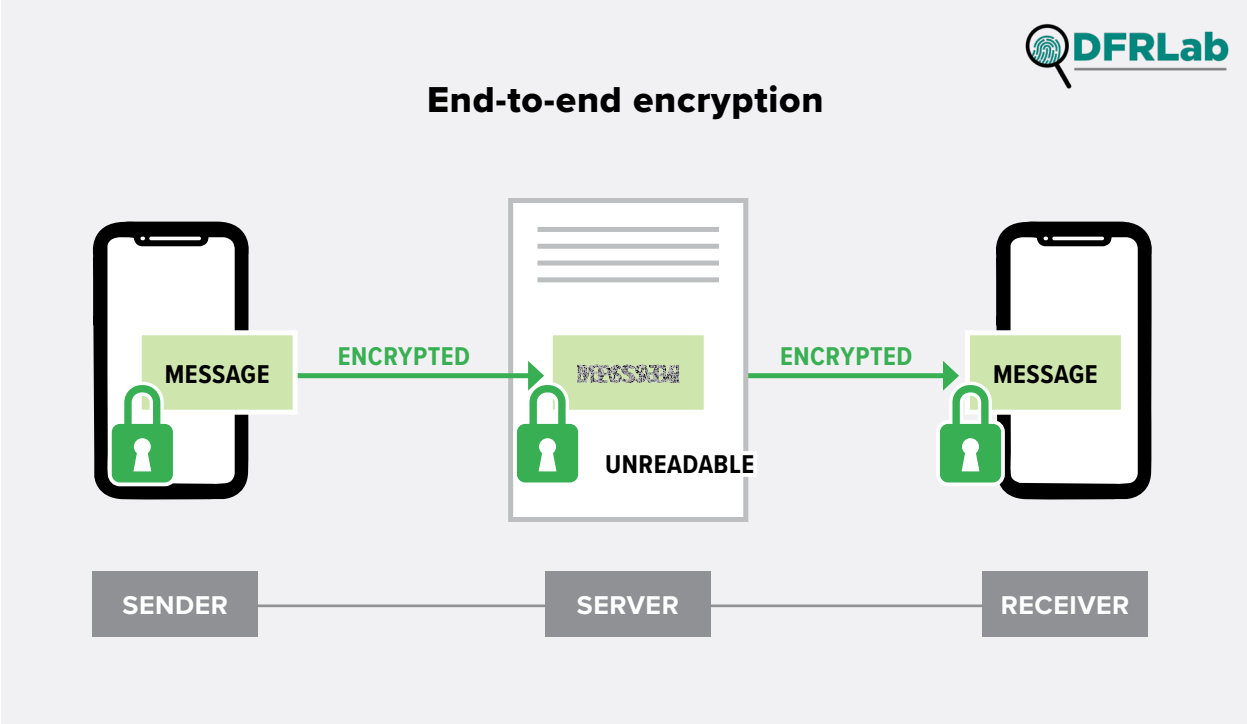


Diagram of mobile messaging app end-to-end-encryption, showing the encryption relay from sender, through the server, and to the receiver. (Source: DFRLab, 2022.)

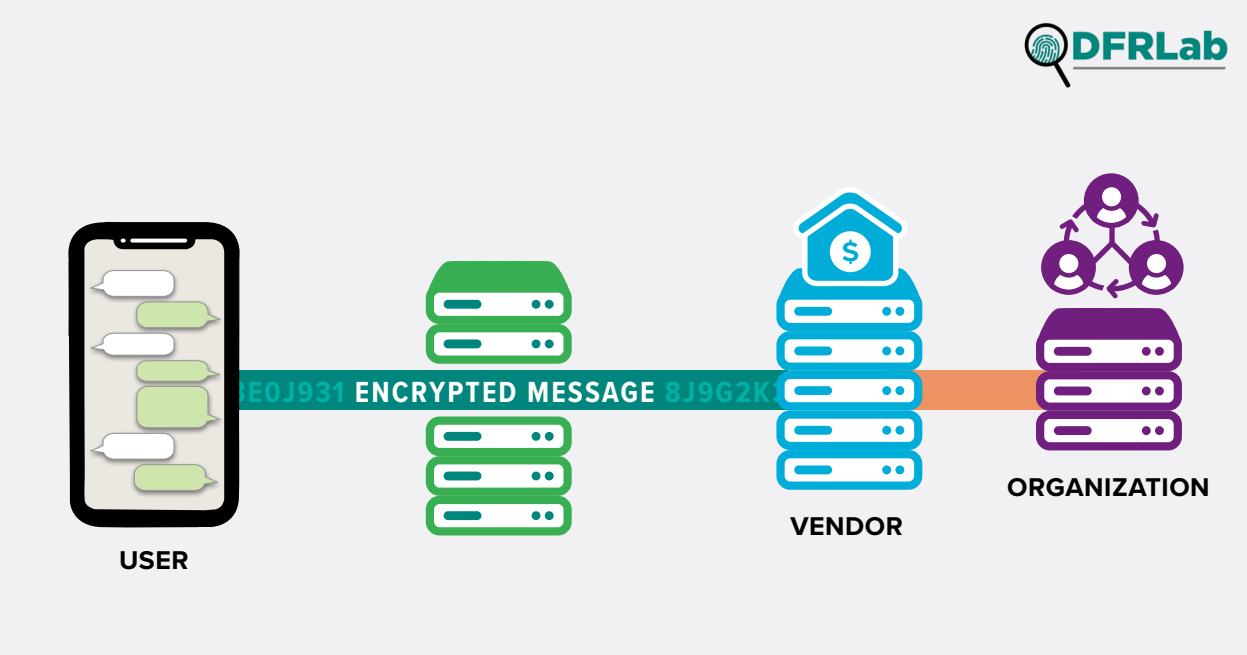


Diagram of the encryption, when a business API account holder delegates operations to a third-party vendor. (Source: DFRLab, 2023, inspired by WhatsApp Encryption Overview, Version 6, Updated November 15, 2021, https://www.whatsapp.com/security/WhatsApp-Security-Whitepaper.pdf.)

While Apple's iMessage offers E2E encryption between iPhone users, its encryption does not work for communicating with somebody using a phone with a different operating system than iOS. iMessage requires user identification, and security experts and digital rights organizations have questioned the security and trustworthiness of Apple's cloud backup due to vulnerability to third-party access, the possibility of backdoors, and the company's past willingness to respond to data requests from law enforcement.[9] In December 2022, Apple announced that the company will implement fully encrypted opt-in server backups to better protect users' photos and chat backups.[10]

Although the integrity of the E2E encryption assumptions may not be technically made vulnerable, there are some caveats regarding what E2E encryption entails when one of the *ends* is not an individual but a business. Large business organizations may provide third parties in their supply chains with access to their conversation logs and in some cases may even delegate the complete operation of their messaging communication to a third party. The individual user interacting with a business account may not realize that all the information exchanged may be shared with several parties. Also, conversation logs between an individual and a business may be kept on organizational computers with varying levels of security.

9 Access Now, "Access Now urges Apple to restore trust by rolling back plans to circumvent end-to-end encryption on its devices," Access Now, August 10, 2021, https://www.accessnow.org/apple-encryption-expanded-protections-children/.

10 Sam Sabin, "Apple to offer encryption on iCloud backups," Axios, December 7, 2022, https://www.axios.com/2022/12/07/apple-encryption-icloud-backups.

Privacy and Security Features for Prevalent Messaging Apps in the United States

App	Registration	User Identity	E2E Encryption
WhatsApp	Phone Number	Phone number	E2E encryption by default; in-server encryption for back-ups.
Telegram	Phone Number	Username	Opt-in E2E encryption using "secret chats"; in-server encryption for private chats; no encryption for public chats.
WeChat	Phone Number	User ID	No E2E encryption; only TLS for transactions.
Line	Phone Number	Username	Optional encryption.
KaTalk	Phone Number	User ID	Optional encryption.
Viber	Phone Number	Phone number	Optional encryption.
Kik	Email	Username	No E2E encryption; only TLS.
Discord	Email	Username	No E2E encryption; only TLS.
Facebook Messenger	Facebook User (Phone Number/Email).	Username	Opt-in E2E encryption.
Wickr	Does not require registration.	User-generated ID	E2E encryption by default.
Wire	Email/Phone number	Public or private key	E2E encryption by default.
Signal	Phone number	Phone number	E2E encryption by default.
iMessage	Apple ID (Phone number/ email).	Username	Encryption only between Apple users; encryption in server backups.
Threema	Does not require registration.	User-generated ID	E2E encryption by default.
Snapchat	Email/phone number	Username	E2E encryption

Table summarizing basic privacy and security features across messaging apps most commonly used in the United States. (Source: DFRLab, 2022.)

Telegram, WhatsApp, and WeChat: Background and Policies

As stated before, the DFRLab grounded this report in observations from three case studies conducted in Telegram, WeChat, and WhatsApp. This section provides general information on each platform and describes the analytical approach for each. We discuss the findings in the later section on issues and trends.

Telegram

Brothers Nikolai and Pavel Durov launched Telegram in 2013. The Durovs had previously developed Russian social network VKontakte, which is now owned by Russian state-owned bank Gazprombank. Telegram's operational center is now located in the United Arab Emirates, while its parent company, Telegram Messenger Inc., is registered in the British Virgin Islands. As Telegram grew out of Russia, even while technically not being housed in the country, its user base understandably comprises a sizable Russian or Russian-speaking contingent, which is also reflected in the high relative volume of Russia-related content on the platform.

In 2018, Telegram was the center of a high-profile conflict with the Russian government over requests for backdoor access to user data. At that time, the app was banned in Russia. Since Russia lifted the ban in 2020, Telegram has once again grown in popular use in the country. Telegram is also popular in several other Eastern Europe and Caucasus countries. Around 2 percent of the total Telegram audience, or approximately ten million users, are estimated to be in the United States. Elsewhere, the Islamic State of Iraq and al-Sham (ISIS), often referred to as the Islamic State group, had used Telegram for recruitment and propaganda in Europe between 2015 and 2018. These antecedents often have led US media to frame Telegram as a hideout for violent extremists, although the app has more than 700 million users worldwide, and it is commonly used to share news by users in Eastern Europe, Asia, and Latin America.

Telegram is a cloud-based messaging app that emphasizes privacy, synchronization across multiple devices (phones, desktop, web version), group interactivity, and a hands-off content policy. The app is not encrypted by default, but users can activate so-called *secret chats*, which are E2E encrypted, do not synchronize across devices, do not allow forwarding, allow self-destructing timing, get deleted for both sides of the conversation when one user deletes them, and are not stored in Telegram servers. Telegram has its own E2E encryption protocol, MTProto Mobile Protocol, which is publicly available for independent audits.[11] Public channels and private nonsecret chats are stored in Telegram servers to allow synchronization across devices. Telegram cloud servers are located in different countries worldwide. These servers are encrypted, but their proprietary software for cloud encryption is not available for audits.

Telegram also offers premium accounts with additional features or better performance for enhancing channels. Premium Telegram accounts are not labeled as such for other users to distinguish them easily.

POLICIES

Telegram's terms of service emphasize privacy protection and its minimal collection of personal data from registered users. Telegram promises a high level of privacy to users opting for secret chats, while its public channels are more akin to a social media platform in terms of who can see, read, and react to the content. Telegram states that the platform does not keep content from secret chats stored on its servers. Due to their encryption, secret chats only can be retrieved in the devices used during the conversation, not even in other devices associated with the same user (e.g., mobile phone app or web version).[12] Nonsecret chats remain available in any device that the user has synchronized. Telegram stores on its cloud servers both the content of nonsecret chats and content shared in public channels, where that data is encrypted

11 Telegram, "Mobile Protocol: Detailed Description," n.d., https://core.telegram.org/mtproto/description.

12 Telegram, "Privacy Policy," August 14, 2018, https://telegram.org/privacy.

with a key controlled by Telegram (rather than stored on the user's device as in secret chats). Thus, Telegram's security and privacy varies depending on the mode of interaction setup (i.e., encrypted "secret" chats, private nonencrypted chats, or public groups).

In terms of acceptable content, Telegram is the most permissive messaging app among the three we analyzed for this report. There are no content policies concerning private chats. There are also no specific policies of acceptable content for premium accounts on the platform that differentiate from regular user accounts. Telegram only prohibits sending spam or scamming other users, promoting violence on publicly viewable Telegram channels or bots, and posting illegal pornographic content on publicly viewable Telegram channels or bots.[13] However, after the Paris terrorist attacks in November 2015, Telegram enforced an ad hoc policy of banning channels of the Islamic State group. Following that policy, Telegram has been reporting daily on the quantity of banned terrorist content via a dedicated channel named ISIS Watch.[14]

In its privacy policy, Telegram indicates that the company would only release a user's IP addresses and phone number to authorities when presented with a court warrant for terrorism-related charges. As of the date of completing this report, Telegram maintains that it has never been compelled to do so. However, German news outlet *Der Spiegel* reported that Telegram has fulfilled a number of data requests from Germany's Federal Criminal Police Office involving terror and child abuse suspects.[15] Likewise, after reaching an agreement with Brazil's Superior Electoral Court in March 2022, Telegram adopted measures for monitoring and fact-checking public channels focused on content related to the 2022 Brazilian presidential elections as a means of mitigating the impact of misinformation on the results.[16]

RESEARCH AND ANALYSIS

To conduct our research on Telegram, we subscribed to ten of the most popular public channels focused on domestic US politics. Then, we employed a snowball strategy in which we expanded our dataset to include other relevant public channels whose content was forwarded by the original ten channels.[17] Using this procedure, we expanded to sixty-two public channels that later served as the seed for requesting, through Telegram's API, channels whose public-facing content was forwarded or shared by the initial set.[18] The final dataset comprised nearly six thousand public Telegram channels or chats, on which we conducted a modularity analysis using Gephi's network analysis software. We ran an unsupervised algorithm that helped identify potential communities of interest based on structural relations among the dataset's channels. Structural relations in this Telegram dataset were driven by channels' quotes, forwarded media, external links, and topical language.[19] This sort of structural analysis removes potential confirmation bias. DFRLab researchers did not arbitrarily place channels in ad hoc communities; instead, the community detection algorithm clustered channels based on their structural properties within the whole dataset. This analysis helped us to understand how channels cluster based on their topics of discussion and their sources.

After the clustering, we conducted content analysis on public posts published between November 2021 and February 2022 to understand the topics and viewpoints present in each of the identified communities.[20] Based

13 Telegram, "Terms of Service," n.d., https://telegram.org/tos.

14 Telegram, ISIS Watch channel (@isiswatch), https://t.me/s/isiswatch; Telegram does not provided detailed data on such banned content.

15 "Telegram hält sich neuerdings an Gesetze, zumindest ein bisschen," *Der Spiegel*, June 3, 2022, https://www.spiegel.de/netzwelt/apps/telegram-gibt-nutzerdaten-an-das-bundeskriminalamt-a-0e4d3fcb-8081-4b87-b062-db412bbc294b.

16 Angelica Mari, "Telegram abides to rules and averts ban in Brazil," ZDNET (business technology news site owned by Red Ventures), March 22, 2022, https://www.zdnet.com/article/telegram-abides-to-rules-and-averts-ban-in-brazil/.

17 The DFRLab has chosen not to include a comprehensive list of the channels in order to avoid the risk of driving traffic or attention to them.

18 API stands for "application programming interface" and refers to an HTTP request to ask an application for machine-readable data in JSON or XML formats. Many, but not all, internet-based applications allow researchers to request data using their APIs. The foundations of what APIs are and how they work are thoroughly explained in: Roy Thomas Fielding, "Architectural Styles and the Design of Network-based Software Architectures," (PhD diss., University of California, Irvine, 2000). For this project, DFRLab Research Associate Esteban Ponce De León developed a script that connects to Telegram's API to request channel data on JSON files. The script is available on GitHub, https://github.com/estebanpdl/telegram-api.

19 Modularity analysis is a method to detect communities in a large network dataset. See Vincent D. Blondel et al., "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment* (2008), 1000, https://doi.org/10.1088/1742-5468/2008/10/P10008.

20 Iria Puyosa and Esteban Ponce de León. "Understanding Telegram's ecosystem of far-right channels in the US," Case Study, Digital Forensic Research Lab (DFRLab), March 23, 2022, https://dfrlab.org/2022/03/23/understanding-telegrams-ecosystem-of-far-right-channels-in-the-us/.

on the topics, language, and perspectives observed in each cluster, we labeled each community with a descriptive name to facilitate explaining to the readers the sort of content found. The labeling attempts to capture commonalities, and some channels may be more representative than others of a particular political identity. Some of the communities identified in the analysis may overlap in their offline political positions. Still, they differentiated in their clustering due to their quotes, linking, and forwarding behavior. Otherwise, groups that are not necessarily in the same specific political community—e.g., Black Live Matters, anarchists, and anti-capitalists—were clustered together by the community detection algorithm likely due to similar topical focus, even when their views varied.

Relatedly, the DFRLab did not analyze individual subscribers to these channels; as such, the clusters should not be understood as representative of the individual personal beliefs of any given channel's subscribers. It is also worth noting that the content shared between channel groups was not necessarily mutually exclusive, nor were the channel members.

WeChat

WeChat is an international version of Chinese company Tencent Holdings' flagship, internal-to-China Weixin app; the apps were launched in 2012 and 2011, respectively. WeChat operates in compliance with the Chinese cybersecurity and national intelligence laws, which mandate platforms hand over information to Chinese intelligence agencies.

The app is all-encompassing and multipurpose. It includes instant messaging as one of its core functionalities, a digital wallet, gaming, public accounts subscriptions, and friends' "Moments" feed (akin to Facebook's newsfeed). Shopping, advertising, and corporate accounts are central to the WeChat business model. WeChat does not offer E2E encryption or cloud encryption.

Besides personal individual accounts, WeChat offers two different types of business accounts: official accounts, which are allowed to push notifications to subscribers, and service accounts for e-commerce. Public accounts may forward content from other official WeChat accounts or link to external sources. Nonsubscribers can read public accounts' posts but are not allowed to react or share. For our research, we analyzed public accounts as nonsubscribers.

POLICIES

In its privacy policy, WeChat acknowledges extensive data collection, including personal data, log data, location, and content shared within the platform's different surfaces ("Profile," "Moments," and "Status").[21] Chat texts are kept on platform servers for three to seventy-two hours, and media is stored for up to 120 hours. Nonetheless, if a user marks a piece of content as a favorite or pins it as a "group notice," that content is stored on the server for an undefined time. Likewise, WeChat may disclose personal data and content to law enforcement[22] and government agencies[23] in the jurisdictions in which the platform operates. In addition, when WeChat users interact with users in China, the platform may collect personal information without seeking consent.

WeChat's terms of service outline the platform's interoperability with Weixin, the licensing of all content shared within the platform to WeChat International or Tencent International Services Europe, the allowance to retain content after an account is deleted, and the allowance to disclose the content to law enforcement and government agencies.[24]

The interoperability with Weixin users is a blanket acceptance of data collection. Users' personal information can be handed to law enforcement agencies (in any jurisdiction in which the platform operates) if related to national security or national defense; to public safety, public health, or major "public interests"; or to a criminal investigation, prosecution, trial, or execution of judicial decisions. User data will also be made available to government or law enforcement agencies if it is necessary to protect someone's life or property; if it is required for contractual matters; if it has been made publicly available in "legitimate news coverage" and governmental announcements; if it is required for ensuring the safe and

21 WeChat, "WeChat Privacy Policy," March 22, 2022, https://www.wechat.com/en/privacy_policy.html.

22 WeChat International, "Law Enforcement Data Request Guidelines," October 15, 2019, https://www.wechat.com/en/law_enforcement_data_request.html.

23 WeChat International, "Governmental Request Policy," August 19, 2021, https://www.wechat.com/en/government_request_policy.html.

24 WeChat, "WeChat—Terms of Service," March 1, 2023, https://www.wechat.com/en/service_terms.html.

stable operation of the provided products or services; if it is necessary for statistical or academic research in the public interest; or under other obligations imposed by laws and regulations.[25] This final rationale in particular operates as a catchall, opening the door for law enforcement to access user information for most any reason; a user may have a lower expectation of privacy on WeChat.

In its acceptable use policy, WeChat acknowledges that the platform deploys automated processes to detect and prevent harmful content.[26] WeChat lists seventy-eight prohibitions that users must abide by, including thirty-five types of content, thirty-five activities or behaviors, transactions over five lines of products, and three categories of users. Prohibited content includes, among others, violent, criminal, or illegal content; threats to public safety; threats to others; terrorism and organized hate; child nudity and exploitation; sexual exploitation of adults; and political promotional content. The platform also prohibits spam, identity misrepresentation, coordinated inauthentic behavior, infringement of intellectual property rights, and any activity that may cause a technical disruption of WeChat. Depending on the jurisdiction, WeChat may prohibit the sales of drugs, including medical or pharmaceutical drugs; ammunition and weapons, including 3D printing items; trade of human organs and blood; alcohol and tobacco; and weight loss products and cosmetic surgery. Finally, WeChat bans people convicted of child abuse or sex offenses, and people less than thirteen years old. There are not specific policies of acceptable content for WeChat public accounts that differentiated from individual users' accounts.

Regarding acceptable content policy, WeChat is the most restrictive messaging app among the three we analyzed for this report. The platform ensures compliance by deploying automated monitoring of content shared in all modes of interactions, including in person-to-person chats. According to our research, out of the three messaging apps, WeChat is the platform that provides the least privacy and the least personal data protection to its users.

RESEARCH AND ANALYSIS

We analyzed a sample of ten official public accounts for the Chinese Students and Scholars Associations (CSSA) affiliated with the US universities with the largest enrollment of Chinese students.[27] The DFRLab collected public CSSA WeChat posts published between October 2018 and February 2022. The final dataset for this study comprised 14,692 posts. We identified a sample of 119 posts conveying political narratives using a logistic regression classifier. Then, we conducted additional content analysis on that sample.[28]

Number of Posts and Reads by University CSSA

CSSA Affiliation	Number of Posts	Reads
Columbia University	1,249	1,551,233
Pennsylvania State University	1,240	821,433
University of Michigan at Ann Arbor	1,860	1,400,978
University of Wisconsin at Madison	2,797	1,987,457
University of California at Berkeley	1,934	1,874,270
Carnegie Mellon University	777	513,716
Purdue University	1,237	690,246
Harvard University	698	1,065,775
University of Maryland at College Park	853	385,120
University of Illinois at Urbana Champaign	2,047	1,347,357

Table summarizing number of posts and number of reads by university CSSA. (Source: DFRLab, 2022.)

25 Weixin, "Privacy Protection Guidelines," January 12, 2022, https://weixin.qq.com/cgi-bin/readtemplate?lang=en_US&t=weixin_agreement&s=privacy&cc=CN.

26 WeChat, "Acceptable Use Policy," March 1, 2023, https://www.wechat.com/en/acceptable_use_policy.html.

27 The universities are Columbia University; Pennsylvania State University; the University of Michigan, Ann Arbor; the University of Wisconsin-Madison; University of California, Berkeley; Carnegie Mellon University; Purdue University; Harvard University; the University of Maryland, College Park; and the University of Illinois Urbana-Champaign.

28 Iria Puyosa, "WeChat channels keep Chinese students in US tied to the motherland," Case Study, Digital Forensic Research Lab (DFRLab), August 31, 2022, https://dfrlab.org/2022/08/31/wechat-channels-keep-chinese-students-in-us-tied-to-the-motherland/.

WhatsApp

WhatsApp is the most popular messaging app worldwide with over two billion users. The app was launched by Jan Koum and Brian Acton (both formerly of Yahoo!) in 2009, and the app grew in popularity after Facebook (now Meta) bought it in 2014.

WhatsApp is E2E encrypted by default, in all its interaction modes (individual-to-individual chats, private groups, public groups, distribution lists, and individual-to-business chats). WhatsApp adopted the Signal-encryption protocol, which is considered the most robust employed in messaging apps and is increasingly becoming the standard for encrypted messaging.[29] The Signal protocol properties offer confidentiality and integrity that enhance security in messaging exchanges.[30]

Beyond individual user accounts, WhatsApp offers two types of business accounts: *small business*, which is similar to a regular user but with a few added features, and *WhatsApp business platforms*, which allow in-app transactions.

POLICIES

WhatsApp offers a high level of security, given its E2E encryption in all its modes of interaction setups (i.e., individual-to-individual, closed private groups, or public groups). Nonetheless, WhatsApp acknowledges that the platform collects and processes extensive metadata on usage, contacts, location, and even media content forwarded or displayed in profile pictures.[31]

According to WhatsApp terms of service, the platform cannot be used to share content that is obscene, defamatory, threatening, intimidating, harassing, hateful, racially or ethnically offensive; encourages violent crimes; endangers or exploits others, especially children; coordinates to cause harm; involves publishing falsehoods, misrepresentations, or misleading statements; or impersonates someone else.[32] All images depicting the sexual exploitation of children (known as child sexual abuse material, or CSAM) and possibly most sexual content shared in public groups falls under prohibited content limits according to WhatsApp's terms of service.[33] The prohibition includes not only messages but also status and profile photos.

However, the fact that conversations and groups are E2E encrypted means that content cannot be automatically scanned or monitored for compliance. In addition, the terms of service prohibit interfering with or disrupting the platform's security, confidentiality, and integrity. This last point implies that backdoor access is ruled out, since E2E encryption guarantees messaging confidentiality and integrity.

Business account holders must adhere to the WhatsApp Commerce Policy, which prohibits transactions with cryptocurrencies, initial currency offerings (ICO), multilevel marketing, payday loans, as well as weight loss products advertised by generating negative self-perception.[34] WhatsApp Commerce Policy also prohibits videos or live shows for adult entertainment, as well as overtly sexualized positioning of products and services.[35] This comprehensive policy eliminates the challenges of identifying the difference between the consensual exchange of sexual content versus sexual content shared in the context of criminal activities such as sexual abuse and human trafficking.[36]

29 For technical analysis of the Signal protocol, see Katriel Cohn-Gordon et al., "A Formal Security Analysis of the Signal Messaging Protocol," *Journal of Cryptology* 33, no. 4 (2020): 1914-1983, https://ieeexplore.ieee.org/document/7961996.

30 Among the most important allowances of the Signal protocol are the following properties: identity key authentication of users sending or receiving messages in a conversation; second pairs of temporary keys are created each time two users exchange messages, making it harder for a third party to gain access to conversations; and plaintexts of messages are inaccessible to third parties, making them useless in proving that a user said anything in a conversation.

31 WhatsApp, "WhatsApp Privacy Policy," January 04, 2021, https://www.whatsapp.com/legal/updates/privacy-policy.

32 WhatsApp, "WhatsApp Terms of Service," January 04, 2021, https://www.whatsapp.com/legal/updates/terms-of-service/?lang=en.

33 WhatsApp's terms of service prohibit sharing content that's illegal, obscene, defamatory, threatening, intimidating, harassing, hateful, racially or ethnically offensive, or instigates or encourages conduct that would be illegal or is otherwise inappropriate. See "How to stay safe on WhatsApp," n.d., https://faq.whatsapp.com/515486185838818/?locale=en_US.

34 WhatsApp, "WhatsApp Commerce Policy," January 15, 2021, https://www.whatsapp.com/legal/commerce-policy.

35 WhatsApp, "WhatsApp Commerce Policy."

36 Investigations on human trafficking indicate that recording sexual content to share online is a common form of sexual exploitation and forced labor. See US Department of State, 2022 Trafficking in Persons Report, July 2022, https://www.state.gov/reports/2022-trafficking-in-persons-report/.

RESEARCH AND ANALYSIS

For our WhatsApp research, we joined ninety-eight public groups via three paths: invitations from administrators who agreed to allow us to join their groups; joining via links posted on Facebook public groups; and joining via links posted to Reddit and other online forums. Overall, we reached two different sets of WhatsApp groups. One set was comprised of Latino-identifying diaspora communicating in Spanish about shared concerns (sixty-six groups). These groups were made up of people originally from Argentina, Colombia, Cuba, Chile, the Dominican Republic, México, Nicaragua, and Venezuela, as well as more general groups of Latinos including US-born Latinos. Groups participants were based in California, the District of Columbia, Connecticut, Florida, Illinois, Louisiana, Massachusetts, Michigan, Minnesota, New Jersey, New York, North Carolina, Ohio, Oregon, Tennessee, Texas, and Washington. For transparency, we communicated in these diaspora groups that we were researching the usage of messaging apps. In most cases, users did not object to the researcher's presence. In two instances, some users asked for the researcher to be banned from the group, after which the administrator removed us. Out of respect for their decision to remove us, we did not use any information potentially gained from these groups within the analysis presented in this report.

The other set was less homogenous, comprised of groups with different topical focuses. Administrators of these groups posted their links to Reddit and Facebook, targeting US-based, English-speaking participants (thirty-two groups). Using a DFRLab account that identified us and stated our research goals, we observed these groups daily over five to seven months (depending on the date of joining the group) to document their communications dynamics and topics of conversations.[37]

37 Iria Puyosa, "Latinos in the US turn to WhatsApp groups for information on the Uvalde shooting," Case Study, Digital Forensic Research Lab (DFRLab), June 3, 2022, https://dfrlab.org/2022/06/03/latinos-in-the-us-turn-to-whatsapp-groups-for-information-on-the-uvalde-shooting/.

Issues and Trends

The DFRLab explored public groups and channels on WhatsApp, Telegram, and WeChat to provide an overview of issues and topical trends, with a particular eye toward identifying opinion-shaping or manipulated content, including disinformation. Although most exchanges within messaging apps are personal, point-to-point conversations, messaging apps are increasingly used for group conversations. This includes private groups, public groups, and interactions among persons and organizations. Focusing specifically on public-facing groups, we chose to look at issues more relevant to public affairs, such as disinformation, foreign influence, electoral politics, and diaspora communities.

Based on our analysis, we identified several trends in the Telegram channels, WeChat accounts, or WhatsApp groups. The DFRLab found that:

- All three messaging apps allow for combining multiple formats (text, audio, images) in the same conversation and linking to content posted in social media platforms to back points or expand the conversation.

- Conversations on these messaging apps intertwined local communities' interests and information related to transnational issues. We observed this trend especially in WhatsApp and WeChat diaspora groups, but it was also present in white supremacist groups on Telegram.

- Diaspora communities rely on WhatsApp and WeChat for mutual support and exchange of resources.

- Misinformation and disinformation about political and health topics were widespread on the public Telegram channels. We also found health-related misinformation in WhatsApp public groups and misleading political narratives on WeChat public accounts.

- Individuals and groups who espouse extremist beliefs, particularly white supremacy, are active on Telegram public channels. We did not observe these types of extremist groups and/or movements in public WhatsApp groups or on public WeChat accounts.

- Public WeChat accounts affiliated with CSSAs were weaponized to foster narratives aligned with the Chinese Communist Party. We also observed pro-Kremlin narratives on public Telegram channels. We did not detect evidence of foreign influence operations on the public WhatsApp groups we analyzed.

- Prominent US Telegram communities and channels identified by DFRLab analysis, especially those who engaged in pro-Trump rhetoric, used public Telegram channels to amplify their political views ahead of the 2022 midterm elections.

- Unsolicited sharing of sexual imagery and content derived from sexual exploitation was prevalent in public WhatsApp groups. We do not observe this sort of content on public WeChat accounts or Telegram channels.

- Some users with business accounts violated WhatsApp policies of acceptable usage.

In the following subsections, the report goes deeper into these trends and patterns.

Multiple Formats and Cross-platform Sharing

Point-to-point messaging apps are often part of multiplatform ecosystems. Most people who use messaging apps also use social media platforms and consume news from digital outlets and traditional sources such as television and radio. In groups and channels, users share content that they created, that they received from other contacts, and that they found on other platforms.

On WhatsApp, typical group conversations involved a series of short texts by which participants reply to each other. Participants also commonly forwarded information they have received in other chats. Besides this sort of WhatsApp-native content, participants also frequently shared content from social media platforms and news outlets. The most commonly shared external content we observed included YouTube videos, links to news sites from both the United States and Latin American countries, Instagram posts, memes, and TikTok videos.

Research on WhatsApp in Latin America, Asia, and Africa has found that anonymous voice messages are used

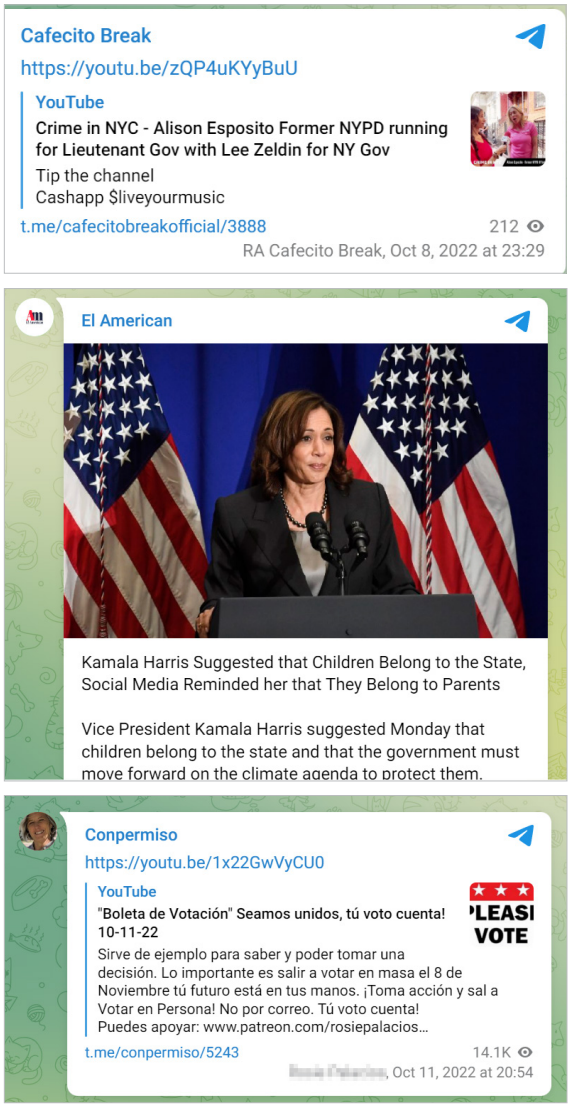
to spread disinformation and to incite violence against vulnerable groups.[38] However, among WhatsApp public groups in the United States, we did not observe such anonymous voice messages. While voice messages did appear in group conversations, they primarily shared lengthy anecdotes, attempted to explain a complicated situation while seeking advice, or expressed emotions regarding personal concerns.

On WeChat, the public CSSA channels (or "official accounts" in the app's terminology) typically publish posts several paragraphs long. Typical posts included illustrations, pictures, emojis, and sometimes Chinese memes; posts also frequently linked to external sources, including the Chinese Embassy in Washington and its US consulates in San Francisco and Chicago, US government websites, and university websites. Some posts referenced Douyin (the Chinese-language precursor to TikTok) and the microblogging platform Weibo.

In our sample of Telegram communities and channels in the United States, Telegram public channels are more commonly used by right-wing than left-wing groups. Besides right-wing channels being more numerous and more active, we observed differentiated patterns in the usage of the available formats between right-wing and left-wing communities. Right-wing channels frequently used videos, pictures, and memes. On Telegram, right-wing channels usually had the chat feature activated, allowing channel subscribers to comment on the posts, which consequently increases participation and engagement. Their texts were usually short and conversational. These channels often forwarded content from other Telegram channels and frequently linked to conservative and far-right news sites, blogs, Twitter accounts, and YouTube videos, as well as references to predominantly right-wing "alternative" social media platforms such as Gab, Odysee, Rumble, Parler, and Truth Social. While some of these groups complained about content moderation and liberal bias on more mainstream platforms, members of these communities preferred to share content originally posted to those platforms.

In contrast, the left-wing channels identified in the DFRLab's research rarely forwarded content from other channels. They shared fewer links to news sites and social media.[39] Left-wing channels were also heavy in text, written more like short op-eds rather than more conversational posts. They used images occasionally, but not as regularly as right-wing channels.

Our dataset included twenty-five Telegram channels directed to the US Latino population, mostly focused on QAnon and COVID-19 conspiracy theories. We found that



Right-wing or conservative Telegram channels targeting Latino users in the United States. (Source, top to bottom: Cafecito Break, October 8, 2022, https://archive.ph/zVMoa; El American, October 12, 2022, https://archive.ph/eamXe; Conpermiso, October 11, 2022, https://archive.ph/vAS4F.)

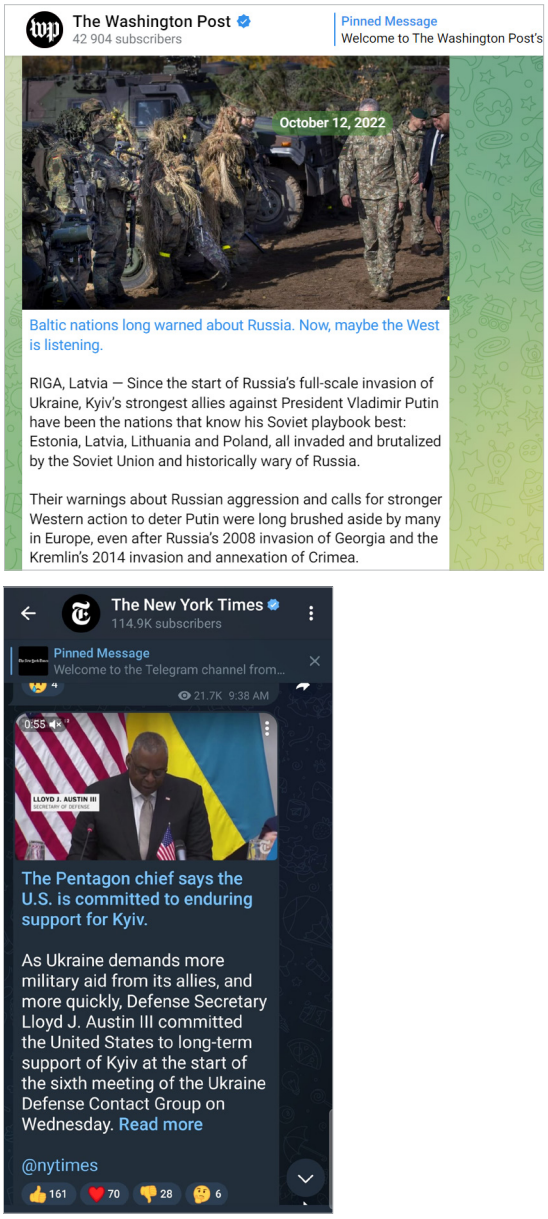
38 Lisa Schirch, ed., *Social Media Impacts on Conflict and Democracy: The Techtonic Shift* (Milton Park, United Kingdom: Routledge, 2021).

39 These patterns are consistent with previous research on differences between far-right and far-left groups using Telegram. See Samantha Walther and Andrew McCoy, "US Extremism on Telegram," *Perspectives on Terrorism* 15, no. 2 (2021): 100-124.

right-wing Telegram channels targeting the Latino population were often linked to streaming shows, YouTube channels, and news sites. Digital outlet El American, which publishes Spanish- and English-language news with a pro-Trump, anti-communist perspective, has a Telegram channel on which it shares links to its own news stories. Cafecito Break, a video stream on Rumble, also maintains a Telegram channel and uses it to reshare its own content targeting US-born young Latinos. Its posts are in both English and Spanish. YouTube show ConPermiso reshares its videos and opens its Telegram channel for ongoing conversation with its audience, which primarily comprises Cuban Americans. Conversely, we did not observe left-wing Telegram channels engaging with the Latino population in the United States.

Telegram public channels are well suited for news distribution. The platform's features facilitate combining formats, embedding videos, sharing external links, quoting and forwarding content from other channels, allowing comments and reactions, and live streaming. All these features are attractive for news channels run by independent journalists, content creators, and small digital news outlets. Small or independent Telegram news channels often request cryptocurrency donations from their subscribers. After Russia's renewed war against Ukraine in 2022, large and reputable media outlets activated Telegram channels to distribute their news headlines directly to audiences on this messaging platform.

For most users, messaging apps are a central part of their daily communication and are integrated into the information ecosystem in which they receive news and discuss issues. Content published on social media or news outlets, including radio and television, may reach audiences in closed messaging apps, though the extent of that reach remains unknown. Similarly, rumors can be amplified through messaging apps. These channels also allow participants to reshare content easily with close personal contacts in a private way, in contrast to social media sharing with more distant acquaintances. Content shareability facilitates building communities and spreading information (and disinformation).



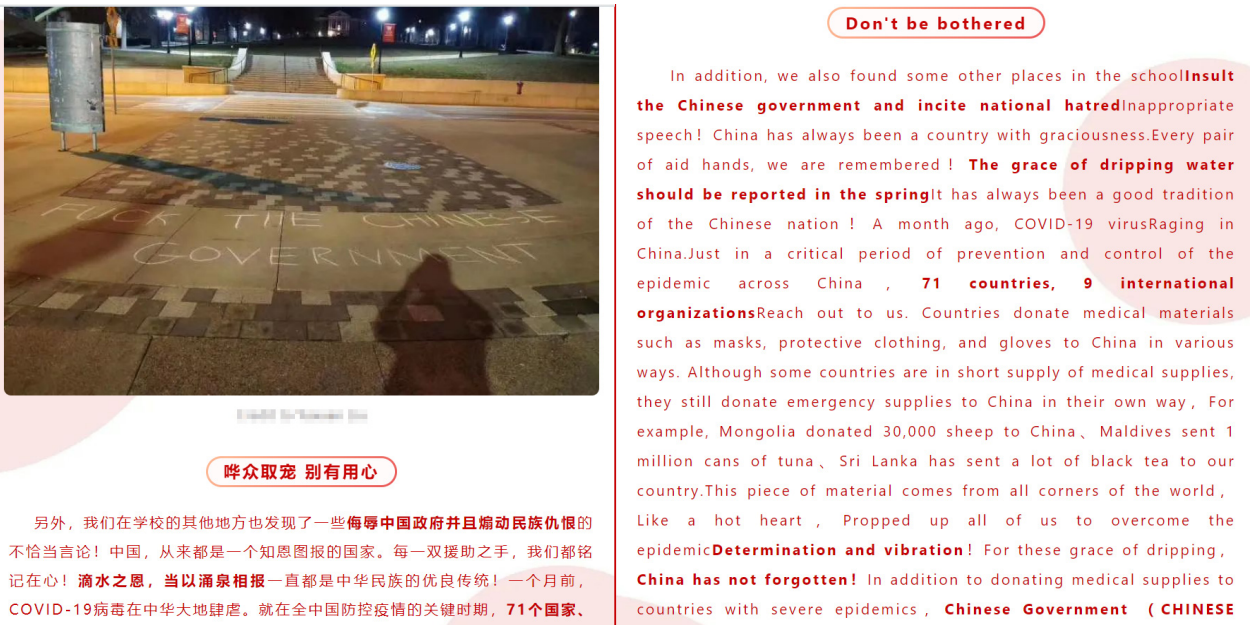
Screencaps of Telegram channels for the Washington Post and The New York Times, which were among the reputable media that opened Telegram channels after the Russian invasion of Ukraine. (Source: The Washington Post, October 12, 2022, https://archive.ph/tD2QU; The New York Times, October 12, 2022, https://archive.ph/mMCD2.)

Transnational Issues and Local Connections

Closed messaging apps enable transnational flows of information. Messaging apps are based on internet networks rather than national mobile phone networks, which substantially decreases the costs of international messaging and voice exchanges. Shared topical interests usually provide the thrust for the content found in messaging app groups and on public channels, facilitating the spread of that content beyond national borders. For instance, the DFRLab observed on Telegram public channels that QAnon conspiracy communities in the United States often connect with like-minded groups in other English-speaking countries, such as Australia, New Zealand, and the United Kingdom, and that US-based anarchist groups shared content from European and Latin American anarchists. Similarly, Latinos shared news from Latin American digital outlets on WhatsApp, and Chinese Americans used WeChat to discuss Chinese news.

Simultaneously, closed messaging apps also reinforce local interactions. The DFRLab observed that white supremacist organizations maintain Telegram channels for state-level chapters in the United States. For example, White Lives Matter and Patriot Front both use Telegram to organize meetups and other local activities. Separately, we also found that Latino migrants often set up state or city-level WhatsApp groups by nationality, such as *Venezuelans in Sugarland-Texas* ("Venezolanos en Sugarland-Texas") or *Mexicans in New Mexico* ("Mexicanos en New Mexico"). Some groups are highly segmented, such as *People from Villavicencio in Los Angeles* ("Gente de Villavicencio en Los Angeles"), which invites people from a specific Colombian region living in a specific US city. We also observed that Chinese college students use CSSA-sponsored public WeChat groups to keep track of campus events.

These interactions between transnational identities and both local and foreign information may significantly affect group participants. Participants appear to define their own identity in relation to groups of people that share the same cultural values. That cultural identity, often based on nationality, ethnicity, ideology, or other shared traits, shapes how participants frame issues of interest. Moreover, participants may use a transnational framing to decide how to react locally. Thus, Chinese students affiliated with campus CSSAs may choose to demonstrate on their campuses against criticisms of China's policies because they consider those criticisms an expression of sinophobia. Similarly, Cuban Americans may vote against



Screencap of a post to the University of Wisconsin CSSA WeChat channel (at left, machine-translated version at right) reacting to graffiti in downtown Madison blaming China for the COVID-19 virus. The incident provoked Chinese students' reactions on the night of March 24, 2020, and subsequent days. (Source: University of Wisconsin CSSA WeChat channel, accessed March 26, 2020, https://archive.ph/r1cJd.)

issues perceived as increased federal government reach in the United States because it reminds them of the Cuban state controlling every facet of people's lives.

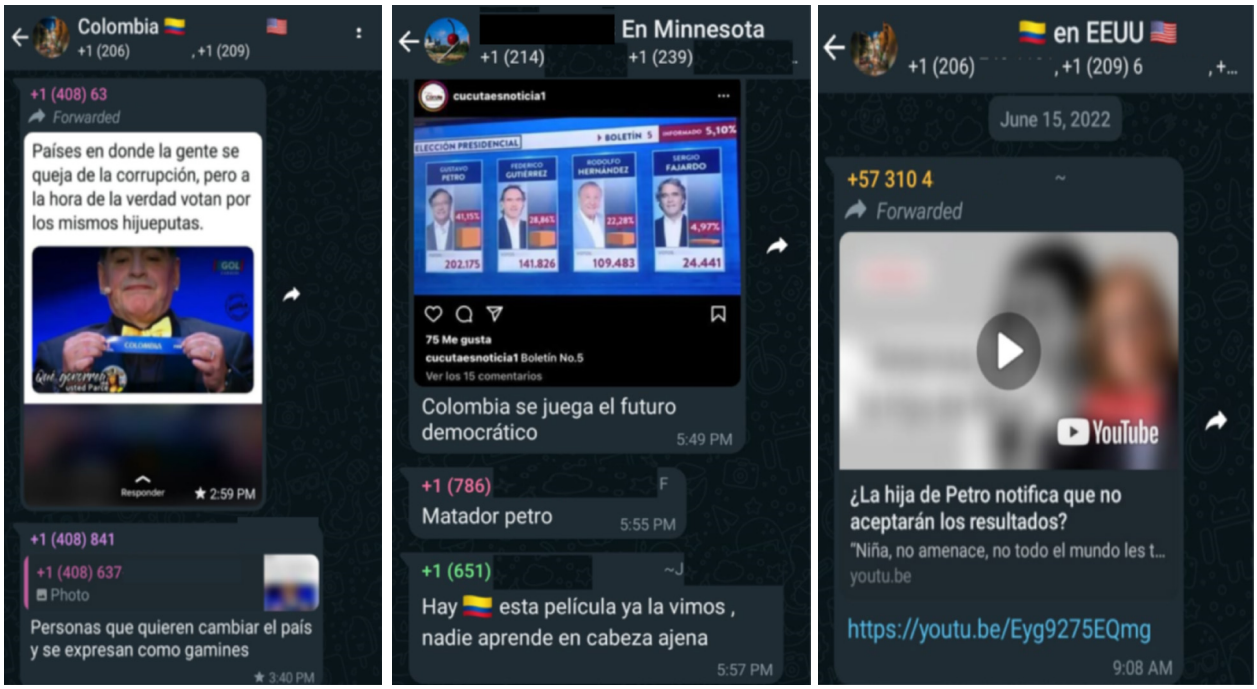
Obviously, these processes of framing and triggers to act also occur in other spaces of interaction, online and offline. The contribution of the messaging apps is to provide immediate feedback confirming that other people in their community are thinking similarly.

Diaspora Groups

Given our interest in understanding the usage of messaging apps by diaspora communities in the United States, we analyzed public conversations focused on issues relevant to immigrant groups. The initial questions shaping the research approach for this project related to how these groups may be used to discuss politics from the countries of origin and how those discussions intersect with US domestic politics. That initial inquiry derived from news coverage of the role of Cuban and Venezuelan politics and the fear of socialism in the 2020 Latino vote in the United States, particularly in Florida.[40]

Our observations of WhatsApp groups indicated that US politics were not a major focus or topic of conversation in the groups. Coverage of major news events around Latin American politics received more attention, as was the case of the Colombian presidential election. Current affairs tangentially discussed in these public groups were usually related to the economy—specifically inflation, gas prices, and employment—or to immigration policies.

It appears that diaspora communities living in the United States prefer messaging apps popular in their countries of origin. One reason for this is that they allow them to keep in touch with their contacts in those countries. While this may apply to every diaspora, it is particularly noticeable within the Chinese diasporas using WeChat, due to its interoperability with Weixin, allowing contacts with mainland China, where other messaging apps are heavily restricted.

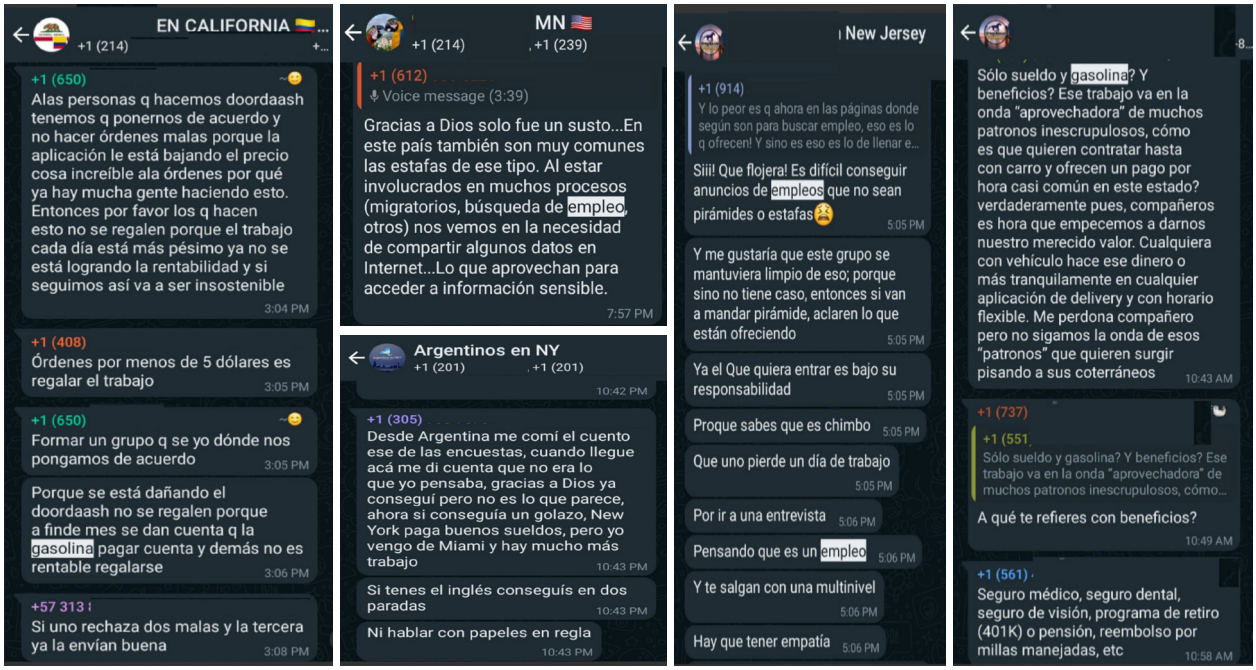


Screencaps from Colombian and Venezuelans WhatsApp groups sharing polarizing content around Colombian Presidential elections. (Source: WhatsApp groups.)

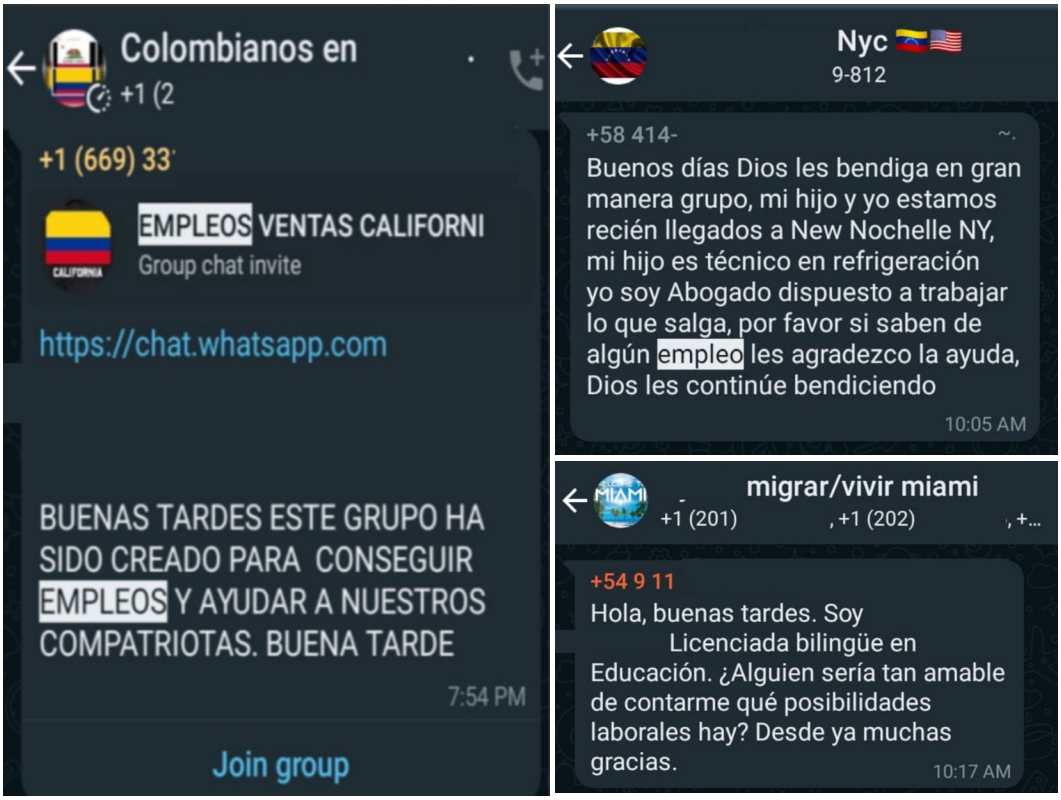
40 Eugene Scott, "Will Painting Democrats as Socialists Help Trump with Latinos?," *Washington Post*, June 26, 2019, https://www.washingtonpost.com/politics/2019/06/26/will-painting-democrats-socialists-help-trump-with-latinos/?itid=lk_inline_manual_19.

We found that immigrants seek to join groups with participants from their country of origin, even if they were not previously acquaintances. We observed Latino immigrants creating Facebook groups to communicate with people of the same origin in their new town, and later setting up a WhatsApp group for more direct communication. Participants leveraged WhatsApp groups to address everyday issues, such as locating groceries stores that sell a particular food popular in their country of origin. We also observed group participants forming social relationships that led to future real-world connections, such as becoming roommates or recommending each other for jobs.

As mentioned previously, the Latino diaspora uses state or city-level WhatsApp groups to build communities and support networks. These Latino groups focus on supporting participants' adjustment to life in the United States and advising those who arrived recently. Most conversations revolved around job offers, housing, and affordable healthcare. Some groups were inclusive of different nationalities of origin, focusing instead on shared lifestyles, such as *Latino Families in South Florida* ("Familias Latinas en el Sur de la Florida"), a channel focused on outdoor family activities, childcare, after-school academic support, and school-related issues. We also found groups focused on sports (soccer, baseball, and softball), which follow international tournaments and organize local amateur games.



Screenshots of WhatsApp posts in which migrants discuss forms of exploitation they experienced in delivery gigs that barely cover transportation costs, through online employment applications with theft of sensitive personal data, fraudulent job offers, pyramid sales schemes, and employers who do not cover medical benefits. (Source: WhatsApp groups.)



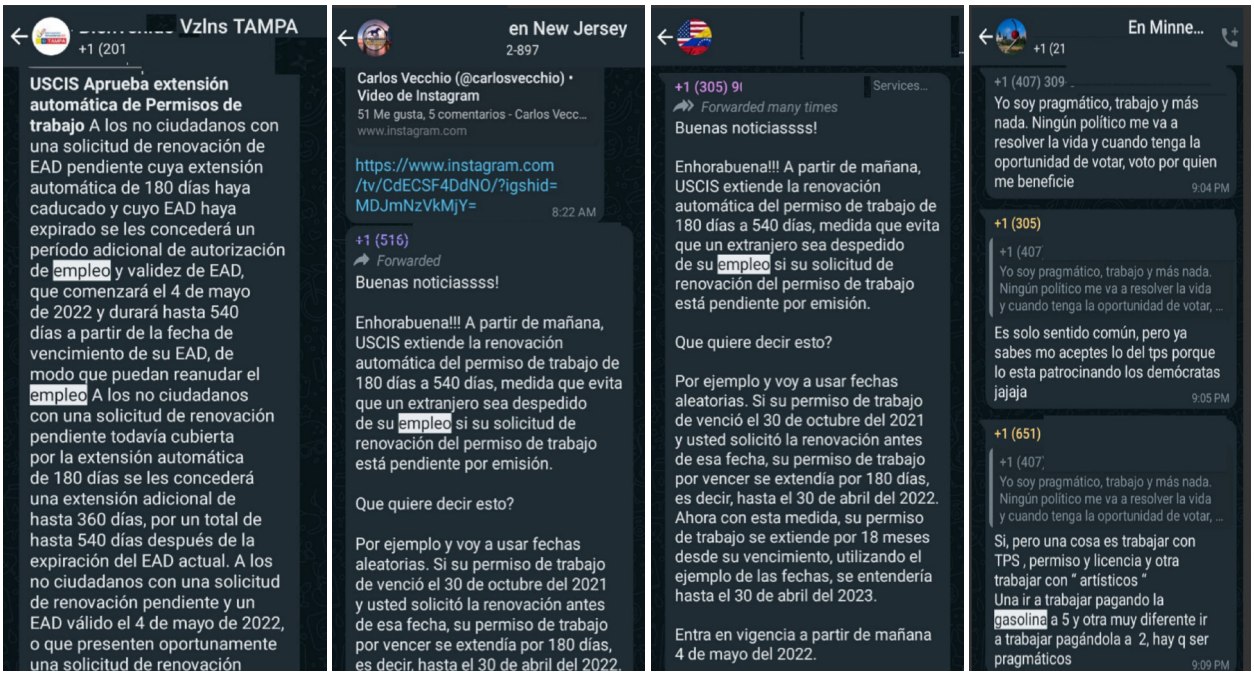
Screenshots of job searches and job offers on migrants' WhatsApp groups. (Source: WhatsApp groups.)

Participants also used WhatsApp groups to crowdsource information about laws, permits, identification procedures, and other bureaucratic processes. Routinely, participants in WhatsApp migrant groups asked for information and advice on filing taxes and obtaining a social security card, employment authorization, or a driver's license. Newcomers asked for help navigating the immigration system, particularly matters related to immigration parole, asylum hearings, work permits, and temporary protected status. When discussing immigration issues, participants commonly referred to lawyers' interactive question-and-answer webcasts held on Instagram and informational materials from nonprofit organizations advocating for migrant rights. Without the guidance provided by more experienced group members, gathering such information could be time-consuming and may require hiring a paralegal or a lawyer, whose professional services may not be affordable for recently arriving immigrants. News related to the designation of temporary protected status for Venezuelans was also discussed in some groups.

Similar processes of crowdsourcing of information may occur regarding news and public affairs. For example, we observed how Latinos living in Texas crowdsourced and translated news from English into Spanish regarding the Uvalde shooting to overcome barriers to accessing timely and reliable news coverage in their maternal language. The case illustrated how messaging apps could enrich the media environment for those with limited English proficiency or simply a preference for getting their news in Spanish.[41]

While identifying links promoting public WhatsApp groups on Facebook, we also found invite links to groups for people planning to migrate to the United States. Some of these groups were managed by people with business accounts with phone numbers from Mexico and Colombia. Most participants in these groups appeared to be originally from Colombia, Venezuela, and Cuba, with some already living in the United States. People in these groups usually discussed the advantages and risks associated with different migration routes, including charter flights from Cuba to Nicaragua or the crossing

41 Puyosa, "Latinos in the US turn to WhatsApp groups."



Screenshots from Venezuelan migrant WhatsApp groups in which users shared news related to extension of the designation of Temporary Protected Status. The extension was approved on May 4, 2022. (Source: WhatsApp groups.)

of the Darién Gap, located on the Isthmus of Panama. Participants also discussed the costs of the journey (*la travesía*), including fees charged by *coyotes* or cartel-approved guides and the bribes migrants needed to pay in Mexico in order to reach the US border. Other conversations focused on what happens when a person crosses the border and voluntarily surrenders to the Border Patrol to ask for asylum. Some US-based participants, who already lived the experience or have relatives who did, describe what it is like to stay in a detention center and how to prepare for an asylum hearing.

Closed messaging groups provide a space for social groups to find shared support. Building trusting relationships between individuals from a shared background is critical for people who may feel vulnerable because of their migration status, limited English proficiency, or perceived cultural differences. Thus, a person who was isolated in a new country can utilize these apps to build a network of local personal ties with people who may provide emotional or material support when in need.

Spread of Misinformation and Disinformation

News stories about disinformation flourishing on messaging apps have caused increasing alarm in the United States, to the point of garnering attention from the US Congress.[42] The list of threats includes COVID-19 misinformation targeting Spanish-speaking WhatsApp users, disinformation attempting to persuade Latinos away from progressive electoral candidates, white supremacist conspiracies spread on Telegram, and Russian narratives attempting to justify the country's invasion of Ukraine.

Media coverage has stressed vulnerability to health misinformation among Spanish-speaking users of WhatsApp. Indeed, early in 2020, misinformation and conspiracy theories about the origins of the COVID-19 pandemic spread among WhatsApp users in the United States and worldwide.[43] In previous research, we observed on WhatsApp false or misleading content related to the pandemic linked to YouTube channels,

42 In 2022, the Senate Subcommittee on Communications, Media, and Broadband, the House Committee on House Administration, and the Congressional Hispanic Caucus, among others, discussed the topic of misinformation or disinformation affecting the Latino population via social media and messaging apps.

43 Luiza Bandeira et al., *Weaponized: How Rumors About Covid-19's Origins Led to a Narrative Arms Race*, Atlantic Council, Digital Forensic Research Lab (DFRLab), February 14, 2021, https://www.atlanticcouncil.org/in-depth-research-reports/report/weaponized-covid-19/.

Instagram accounts, and Facebook pages.[44] In some cases, misinformation circulating in Spanish through WhatsApp text chains could be traced to narratives that had previously circulated through social media in English.[45] Indeed, similar conspiracy theories and disinformation spread across countries worldwide in different languages through Facebook, YouTube, and other social media platforms. Other misleading or manipulated content came from Latin American sources, particularly YouTube personalities. In our WhatsApp analysis, we found a few groups devoted to weight loss that posted misleading claims of keto diet allegedly curing diabetes.

Nevertheless, we did not find any public WhatsApp groups exclusively devoted to spreading misinformation or conspiracies. WhatsApp group participants occasionally shared rumors and misleading or propagandistic content, but that sort of content was not the primary driver of conversations. Rumors referring to US Immigration and Customs Enforcement (ICE) officers placing food delivery orders to capture undocumented migrants doing that job appeared several times in groups of people based in California and Texas. Other participants responded to these rumors by asking the spreader not to post alarming and unconfirmed rumors. Some argued that, if ICE were using that tactic, the media would cover it in their news. Participants in WhatsApp groups often challenged deceptive or inaccurate content by posting links to content with different or better-explained information.

Occasionally, WhatsApp group members reacted to hyperpartisan or biased content and pointed out that it was politically motivated. In several cases, group administrators reminded participants not to share off-topic content and to abide by group rules to avoid discussing politics. That was a common occurrence in the Latino migrant groups in the days before the contested Colombian presidential elections in 2022.

Measures such as limiting the number of recipients a message can be forwarded to, as WhatsApp has been doing since 2018, may not stop the viral propagation of content, but they do add friction to shareability, slowing its spread. This added friction may discourage regular users from forwarding content they receive. However, individuals working on information or disinformation campaigns can easily bypass these measures by arranging their

Screenshots of off-topic political content and sales announcements in WhatsApp groups that provoked reminders from administrators and other group members. (Source: WhatsApp groups.)

44 I. Puyosa et al., *Information Disorders Propagated in Venezuela via WhatsApp and Social Media amid the COVID-19 pandemic* (Caracas: ININCO-Universidad Central de Venezuela-Venezuela Inteligente, 2021), https://covid.infodesorden.org/reporte/.

45 Cristina Tardáguila, "Desinformación for export: cómo contenidos falsos generados en los Estados Unidos llegan a América Latina," Ecuador Chequea (data-checking portal), August 16, 2021, http://www.ecuadorchequea.com/desinformacion-for-export-como-contenidos-falsos-generados-en-los-estados-unidos-llegan-a-america-latina/.

target recipients in distribution lists or groups. This way, a campaign may reach thousands of recipients within a few minutes. The spread of information on WhatsApp can be accelerated by the recent expansion of group size to 1,024 members.[46] This reach can be exponential, as Meta also introduced "communities" when they announced the group size expansion; communities allow a user to send the same message to multiple groups with a single click.[47]

The DFRLab identified public Telegram channels purposely devoted to creating and spreading disinformation and conspiracy theories (e.g., 148 channels devoted to QAnon and 1,217 spreading COVID-19 misinformation) and inciting race- or gender-based hate speech (e.g., 148 white supremacist channels and 196 channels devoted to "shitposting"[48] and misogynistic/racist memes). These channels counted on a base of subscribers actively looking for content or narratives that fit their beliefs.

During the COVID-19 pandemic, anti-vaccine Telegram channels were a gateway for selling fake vaccine certificates.[49] This trend appears to have been more common in Europe and South Asia.[50] However, the DFRLab found offers of fake vaccine certificates in the US anti-vaccine communities on Telegram.

Conspiracy Telegram channels are numerous enough to be specialized in different topics. In our analysis of nearly six thousand public Telegram channels, we observed three differentiated conspiracy communities: QAnon, anti-vaccine, and COVID-19 origin. Misinformation on the COVID-19 pandemic and the effects of the vaccines was rampant on Telegram during the study. Several chats were set to criticize vaccine mandates and to discuss alleged vaccine side effects. A claim repeatedly amplified on anti-vaccine Telegram channels erroneously asserted that vaccinated people are more likely to get sick than unvaccinated. There also is an active group of channels devoted to QAnon conspiracy theories and the "awakening" against the so-called "deep state," including channels amplifying these conspiracies for Spanish-speaking audiences.

Telegram allows channel verification for politicians and media personalities who can provide reference to verified accounts on two other platforms. However, most channels that we encountered in our research were unverified. Since Telegram users can choose any username they want, it is common to find channels appropriating the identity of a famous person or associating themselves with such a person to attract followers. It is hard to assess whether subscribers understand that these channels are not genuinely linked to the person from whom the name is taken. We found hundreds of accounts using Trump as part of the username and a few dozen using John F. Kennedy, Jr., a central figure in QAnon conspiracies purporting that the son of the former president faked his death and will return to reveal the manipulations of the "deep state."

In our research, we did not find outright disinformation spread on the CSSA groups observed on WeChat. Nonetheless, the DFRLab found ideologically motivated disinformation on WeChat more broadly, primarily related to Russia's war in Ukraine. Most of this political disinformation originated with official accounts of Chinese news outlets, some associated with The People's Daily, a news outlet owned by the Central Committee of the Chinese Communist Party (CCP).[51]

The DFRLab found that misinformation and disinformation circulated widely in public Telegram channels. Observed dynamics indicated that some features of Telegram public channels may exacerbate the spread of disinformation, such as large group sizes, lack of channel administrators' identity verification, and the ability to reinforce narratives by easily forwarding and quoting content from other channels.

46 "Communities Now Available," WhatsApp blog post, November 3, 2022, https://blog.whatsapp.com/communities-now-available.

47 "How to Create a Community," WhatsApp, https://faq.whatsapp.com/438859978317289/?cms_platform=web.

48 Shitposting is understood in cyberculture as posting on social media content that does not add any informational or artistic value to the public conversation but instead attempts to derail public debates with distasteful jokes. Others use the terms "trashposting" or "trash talk."

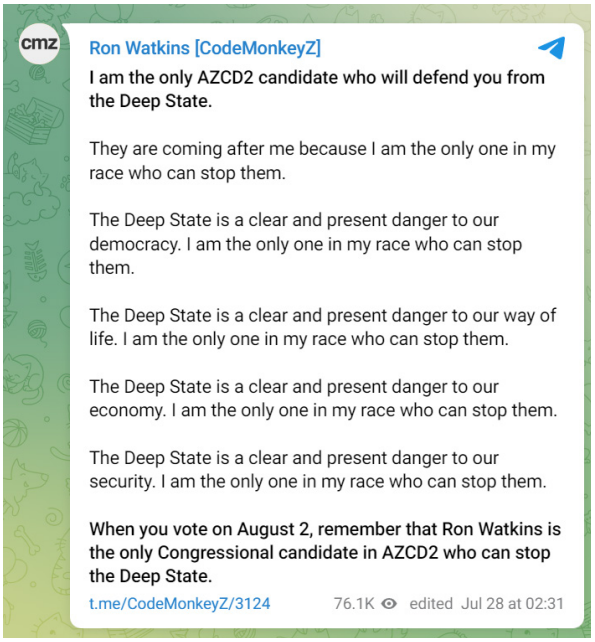
49 For the purposes of this report, the identified channels were "anti-vaccine" in that they pushed narratives that vaccines are intended to induce forced sterilization, undertake mass extermination, or instill mind control. They were not "vaccine hesitant," which would imply general unease or suspicion about the potential side effects of vaccines.

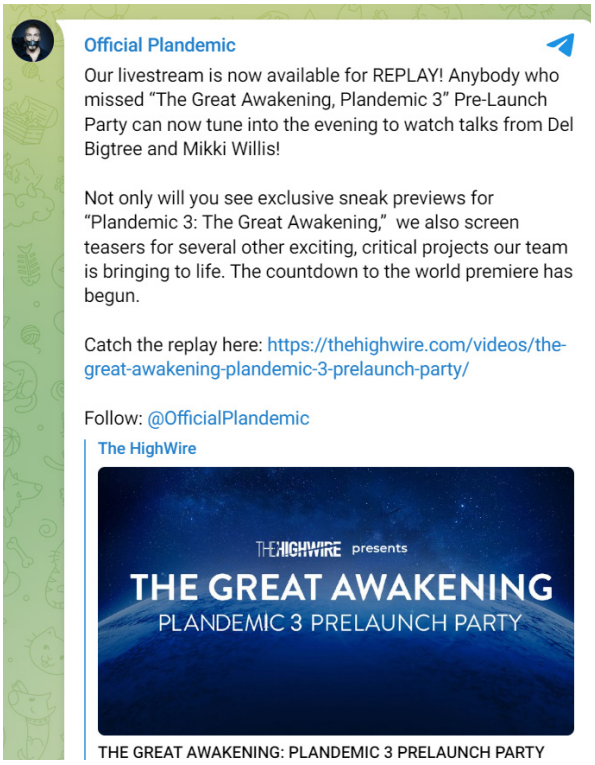
50 Roman Osadchuk, "Scammers use Telegram and Facebook ads to sell fake COVID certificates in Ukraine," DFRLab, December 16, 2021, https://dfrlab.org/2021/12/16/scammers-use-telegram-and-facebook-ads-to-sell-fake-covid-certificates-in-ukraine/.

51 Some recent examples of misinformation and disinformation on WeChat, all of which focus on Russia's ongoing war in Ukraine, can be seen at Weixin, https://mp.weixin.qq.com/s/K9bhifY1BkHcNkOrQa0jOw; https://mp.weixin.qq.com/s/28MnLoibBNVp9hqOEj1Akw; and https://mp.weixin.qq.com/s/T886gV1lm_uWKUynNDkg1Q.

Forwarding and pushing content from other channels also is a feature of WeChat public accounts, though account administrators must provide real identification on the platform. Mandatory identification may restrict misinformation and disinformation to only content aligned with the CCP viewpoints.

On WhatsApp, smaller group sizes and user identification may help to contain the spread of misinformation and disinformation in groups. Closed groups allow administrators to set rules and participants to counter misleading content. However, group participants will not always have the knowledge, interest, or time to verify the accuracy of the content shared in their groups. Besides, the easy resharing of social media content and the enormous transnationally connected user base enable the continuous flow of misinformation and disinformation.





Screenshots of posts to Telegram channels appearing to spread conspiracy theories, including one portraying itself as the late John F. Kennedy, Jr., a central figure in QAnon conspiracies. (Source: "John F. Kennedy Jr." (@John_F_Kennedy_Jr), March 15, 2022; Ron Watkins (CodeMonkeyZ), July 27, 2022; and Official Plandemic (@OfficialPlandemic), September 20, 2022.)

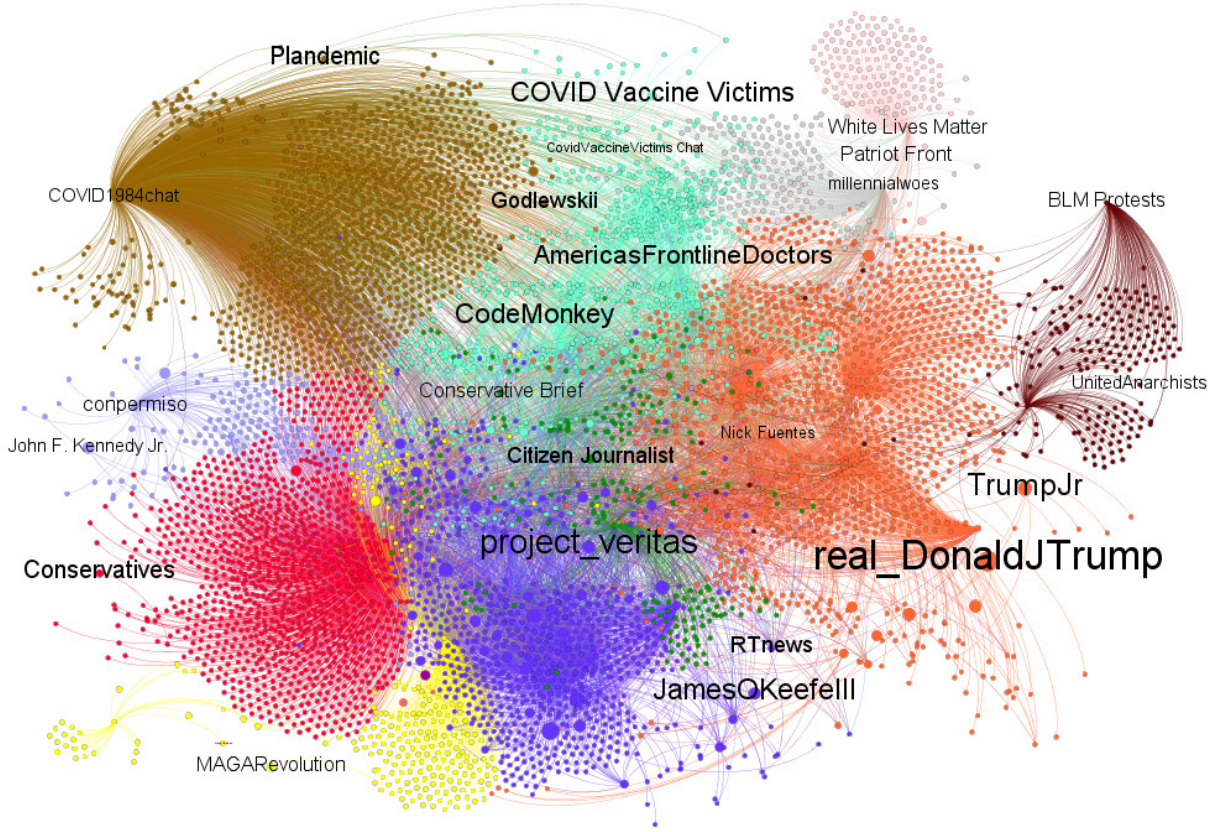
.

Telegram and the Far Right

Telegram experienced a surge in its US user base in early 2021, as right-wing influencers migrated to the messaging platform following Twitter bans related to the 2020 election fraud allegations and the assault on the US Capitol on January 6, 2021.[52]

The interactive features of Telegram public groups favor building connections among groups, since users can easily react and respond to specific posts, as well as quote and reshare to other channels. Right-wing communities appear to be taking advantage of this messaging app for exchanging information and sharing organizing resources.

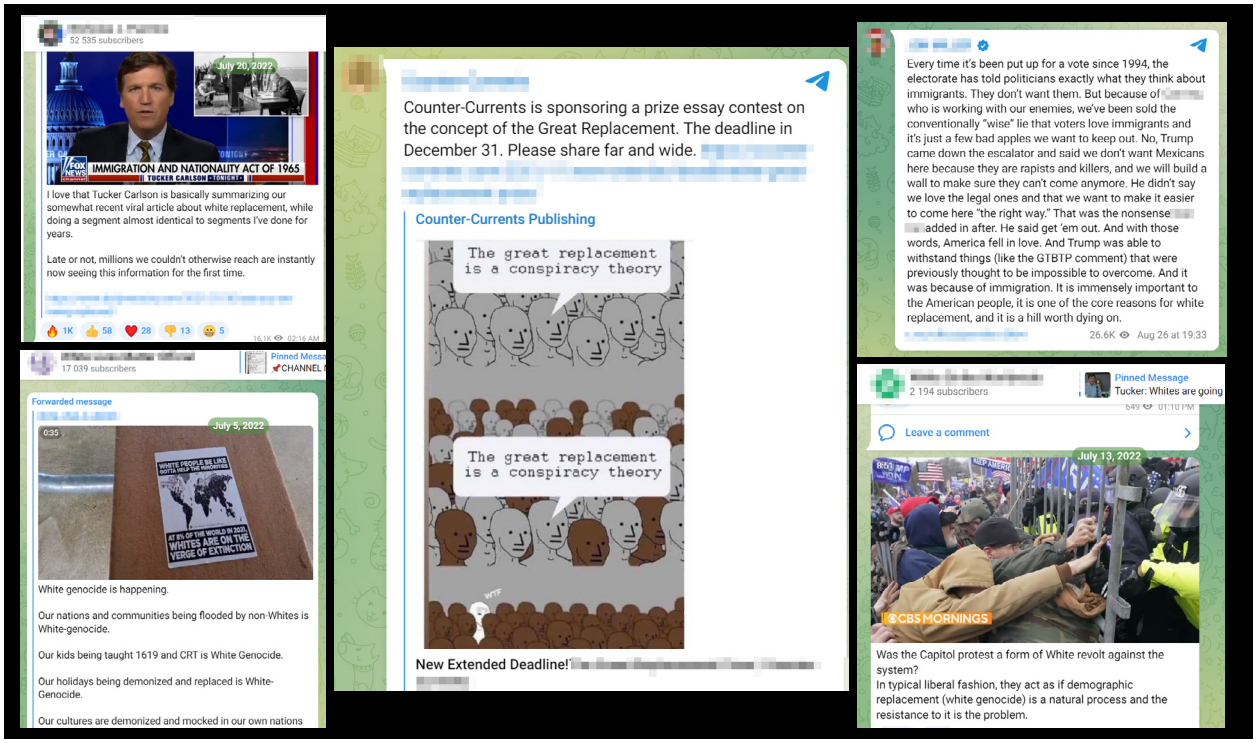
During our analysis, we identified nearly four thousand public Telegram channels associated with right-wing US politics.[53] There is a wide spectrum of users and groups representing the political right in Telegram, from traditional conservative groups to white supremacists and neo-Nazis. More than one thousand Telegram channels explicitly supported Trump (i.e., they were ideologically motivated) or utilized Trump-adjacent themes to drive website traffic and increase ad revenue (i.e., economically motivated).



Network map showing US political communities on Telegram, as of February 2022. The colors represent different communities and each individual point represents a channel. The edges (i.e., lines) connecting the channels are quotes or forwards, and the density of the lines indicates clustering due to the volume of those connections. For legibility, only the most influential channels in each community are labeled. (Source: Iria Puyosa via Gephi, 2022.)

52 Pavel Durov, "Durov's Channel," January 18, 2021, https://t.me/durov/149. For more on how Telegram was used in the United States following the January 6 insurrection, see Jared Holt, "After the Insurrection: How Domestic Extremists Adapted and Evolved After the January 6 US Capitol Attack," Digital Forensic Research Lab (DFRLab), January 2022, https://www.atlanticcouncil.org/wp-content/uploads/2022/01/After-the-Insurrection.pdf.

53 Puyosa and Ponce de León, "Understanding Telegram's ecosystem of far-right channels."



Screenshots of example posts to some of the Telegram channels identified in the course of this research. (Source: Channel names obscured to avoid risk of amplification, post dates ranging November 26, 2021, to August 26, 2022.)

The DFRLab also identified over a hundred white supremacist public Telegram channels in which participants engaged in conversations about the alleged dangers of a multiracial society. Participants in these chats considered a growing Latino population and more politically engaged Black Americans as threats to the white population. The most extreme white supremacists posted content arguing that marriage and procreation between people from different ethnicities or races were forms of "white replacement" or "white genocide."[54]

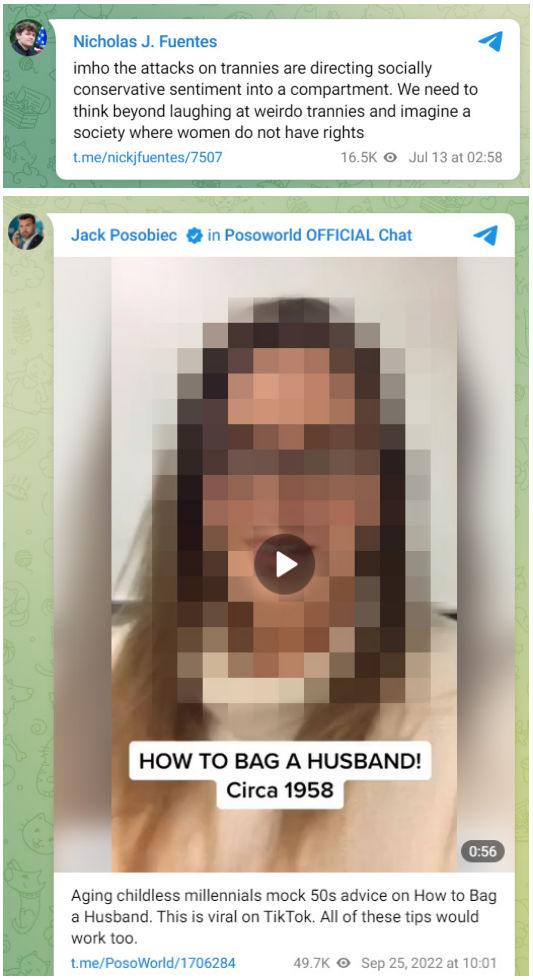
We also identified channels expressing misogynistic and anti-LGBTQ+ views. There are channels mainly devoted to content expressing these viewpoints, while other channels that engage in more broad political discussion sometimes also push misogynistic and anti-LGBTQ+ views. In some posts and conversations, women's civil and political rights were presented as threats to male dominance in public life. Moreover, the recognition of civil rights for LGBTQ+ people was sometimes regarded as a supposed attack on the traditional values in the West.

It is worth emphasizing that Telegram channels are as publicly accessible as many social media platforms are, in that you simply need to register a username in order to join channels and engage with the material or users. So, it

54 For deeper insight into the potential real-world impact of such conspiracy theories, see Matthew Kriner, Meghan Conroy, Alex Newhouse, and Jonathan Lewis, "Understanding Accelerationist Narratives: The Great Replacement Theory," Global Network on Extremism & Technology, May 30, 2022, https://gnet-research.org/2022/05/30/understanding-accelerationist-narratives-the-great-replacement-theory/.

is a reasonable assumption that the individuals or organizations running these channels are seeking engagement, base building, and amplification. White supremacists and other groups in the United States are taking advantage of Telegram's channels to seed extremist views that would not be allowed on some social media platforms.

The DFRLab did not observe similar far-right public groups on WhatsApp. However, we observed some racist, misogynistic, and anti-LGBTQ+ content circulating under the guise of memes or TikTok videos.



Screenshots of transphobic and misogynistic posts to Telegram channels for Nick J. Fuentes and Jack Posobiec. (Source: Nick J. Fuentes, July 13, 2022; Jack Posobiec, September 25, 2022.)

Instrumentalization for Foreign Influence Campaigns

The DFRLab found that messaging apps are yet another digital space for deploying foreign influence propaganda, along with social media and state-affiliated media.

We found that WeChat, for example, is a privileged channel for distributing official Chinese narratives to the Chinese American diaspora, as well as Chinese diasporas in other countries.[55] The DFRLab found that CSSA WeChat groups reinforce pro-Chinese government narratives and even facilitate Chinese student mobilizations in the United States and other countries.[56] In our case study on CSSAs' public WeChat accounts, we identified how these accounts (akin to a public Telegram channel) supported the spread of China's preferred political narratives targeting its diaspora in the United States. The CSSA activities fall under the scope of the Overseas Chinese Affairs Office of the State Council, within the United Front Work Department (UFWD), the entity responsible for spreading CCP propaganda abroad. The DFRLab observed that CSSA accounts supported major CCP political talking points such as the "One China" principle (一个中国原则) concerning Taiwan and the "One Country, Two Systems" policy (一国两制) regarding Hong Kong. Another topic observed during the course of the research was the "great rejuvenation" (伟大复兴),[57] Chinese President Xi Jinping's vision for China's future of economic growth, growing military power, and expanded welfare under the CCP. Also recurrent is the contrast between the "violent and unsafe" United States and the "stable and secure" China. Otherwise, human rights issues such as the forced displacement and cultural erasure of ethnic Uyghurs are unspoken in WeChat CSSA channels.[58]

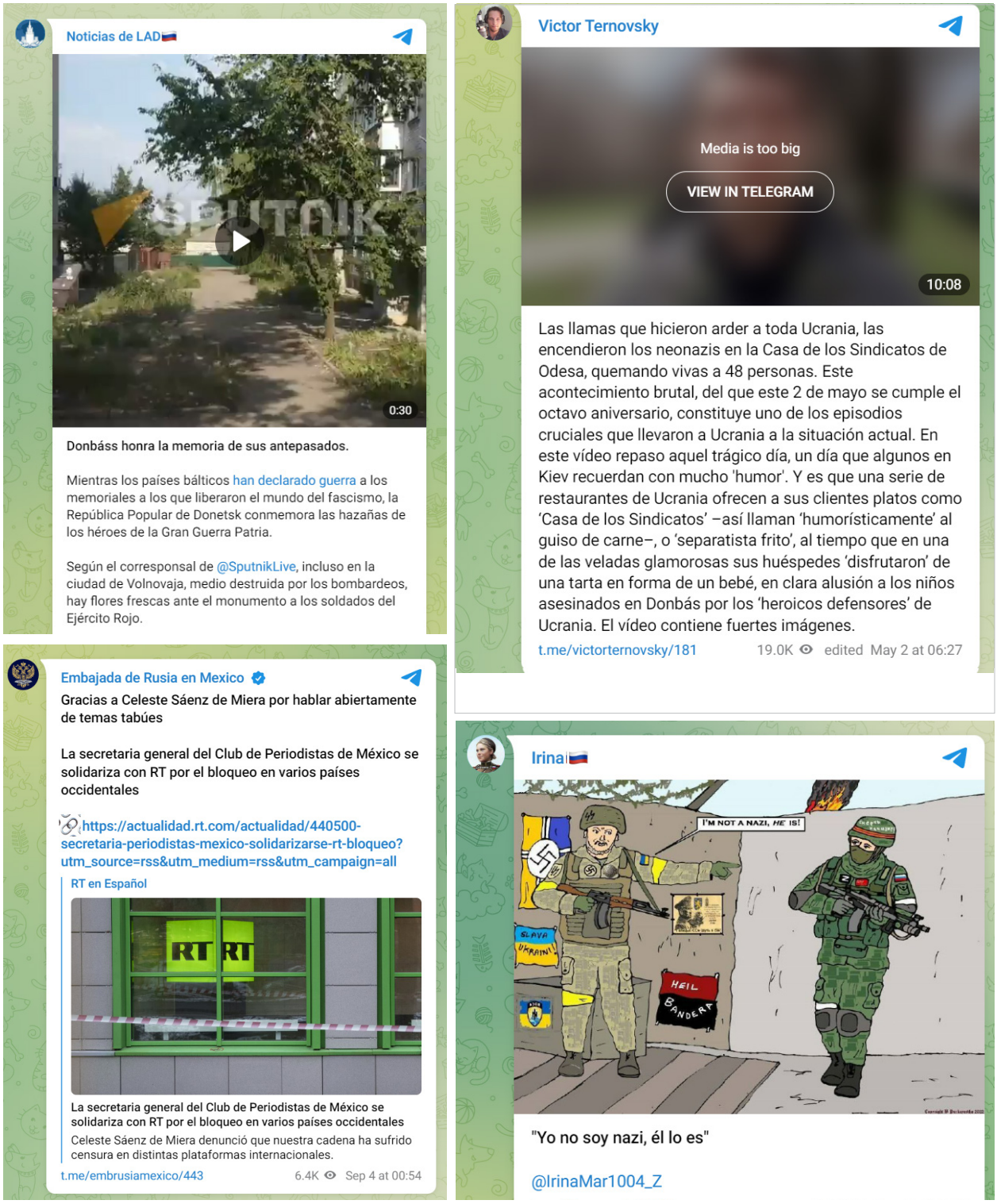
Similarly, the DFRLab found that Russian government and proxies employed Telegram channels to target audiences in the United States, both in English and Spanish. The Russian Minister of Foreign Affairs, for example, organized a vast campaign to amplify Russian narratives on Telegram. After RT News and RT en español were banned following sanctions against Russia, the Latin American Department of the Ministry of Foreign Affairs boosted other Russian government-affiliated Spanish-language

55 DFRLab, "How CSSAs reinforce official narratives to expat Chinese students on WeChat," August 31, 2022, https://dfrlab.org/2022/08/31/how-cssas-reinforce-official-narratives-to-expat-chinese-students-on-wechat/.

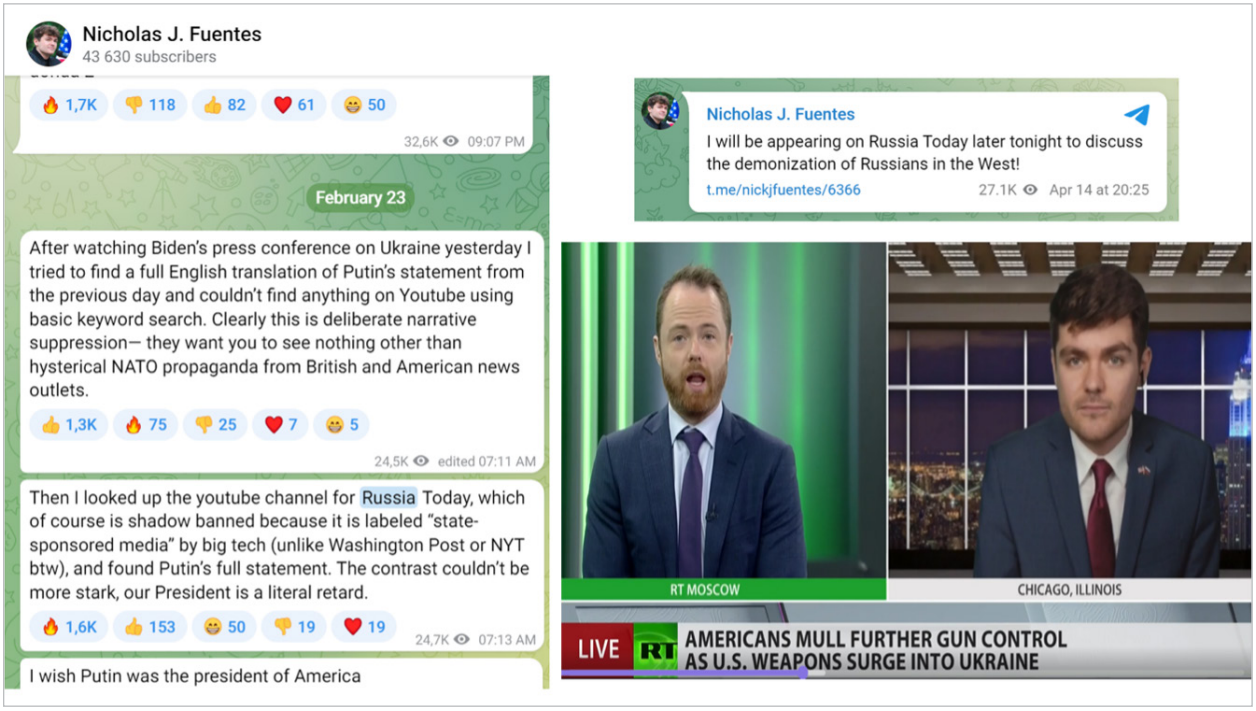
56 DFRLab, "How CSSAs reinforce official narratives."

57 DFRLab and the Snowcroft Center for Strategy and Security, *Descendants of the Dragon*, Atlantic Council, August 17, 2020, https://www.atlanticcouncil.org/wp-content/uploads/2020/12/China-Diaspora-FINAL-1.pdf.

58 Here, we summarize the main topics found in this case study. For more details, see Puyosa, "WeChat channels keep Chinese students."



Screenshots of Telegram channels echoing Russian narratives regarding the war in Ukraine and targeting Spanish-speaking audiences. (Source: Noticias de LAD, accessed September 5, 2022, https://archive.ph/bPBfX#selection-137.0-137.15; Embajada de Rusia en México, September 4, 2022, https://archive.ph/R9vQp; Victor Ternovsky, May 2, 2022, https://archive.ph/tZ3jJ; and Irina (@Irinamar_Z), September 7, 2022, https://archive.ph/QUP5T.)



Screenshots of right-wing influencer Nick Fuentes's Telegram channel, showing Fuentes rooting for Putin and accusing media and tech companies of demonizing Russia. (Source, left to right: Nicholas J. Fuentes, accessed February 23, 2022; Nicholas J. Fuentes, accessed April 15, 2022.)

channels, such as Actualidad RT, RT Latinoamérica, RT Ultima Hora, and Noticias LAD, to continue to spread its preferred narratives on Telegram. The first three are now also banned in the United States. However, Noticias LAD was still accessible at the time of writing this report. This channel shared content from Russian embassies in Latin America and Russian state media. The same content was later amplified by a network of Russian journalists and commentators targeting Spanish speakers in the United States, Latin America, and Spain. These Telegram channels focused on justifying Russia's invasion of Ukraine and pushing anti-Western narratives. Most of the channels only received relatively low engagement, as few of their posts generated more than a thousand reactions or more than a hundred comments. Nonetheless, the amplification network also included nearly a hundred Telegram channels linked to left-wing organizations in Argentina, Colombia, Cuba, Mexico, Uruguay, Venezuela, and Spain. These channels alternated pro-Russian narratives with content related to Latin American politics, supporting the views of the authoritarian left in the region.

We observed similar pro-Russian narratives amplified in English by MAGA profiteering channels and white supremacist channels, though there was no evidence of any formal connection between the Russian government and its amplifiers. That said, the cross-amplification of the pro-Russian narratives was beneficial for engagement to all entities. For instance, America First host and right-wing influencer Nick Fuentes rooted for Putin on his Telegram channel from the outset of Russia's invasion of Ukraine. Fuentes became a regular guest on RT TV broadcasts. Thus, despite banning official Russian media in the United States, Russia managed to instrumentalize Telegram for spreading its propaganda in English and Spanish. Moreover, Russia customized its messaging to attract both left-wing and right-wing audiences; for instance, left-wing Latin America outlets maintain that Russia is defending itself from US imperialistic arm NATO and right-wing outlets also echo the idea that the war originated in NATO expansionism.

It seems that the Russian and Chinese governments have taken advantage of Telegram and WeChat, respectively, to leverage these apps to imbue audiences with false or distorted narratives, with the goal of affecting political views in the United States and other countries. This is also echoed in that Telegram and WeChat were

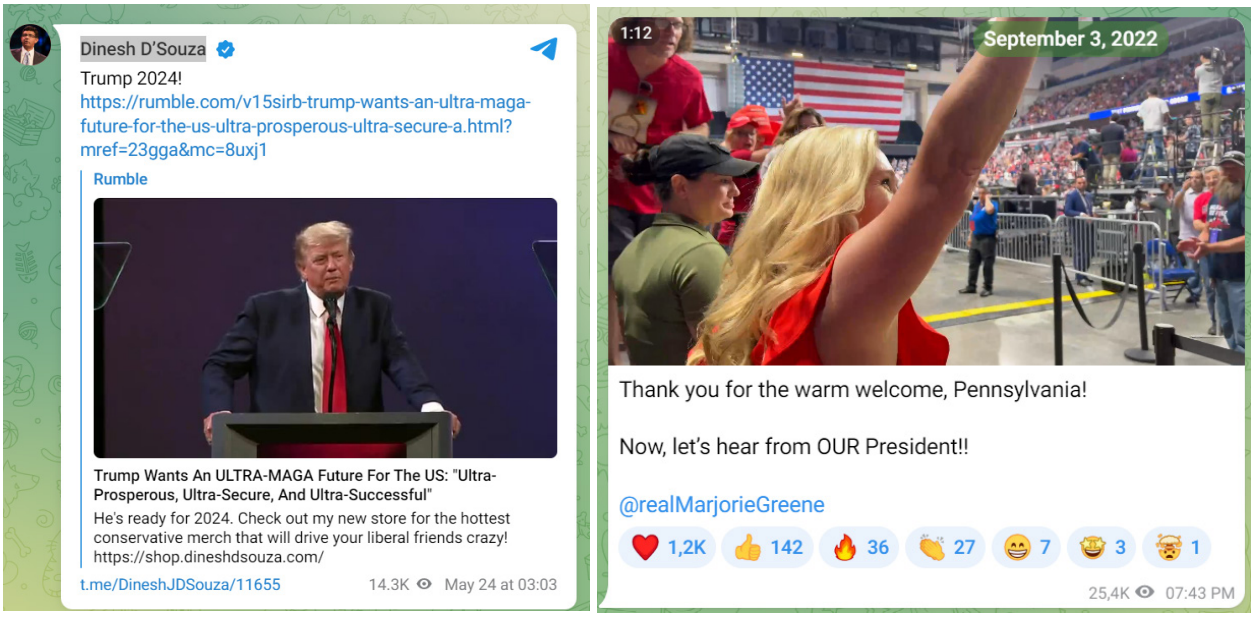
developed and initially headquartered in Russia and China, respectively, and thus pro-Russian and pro-Chinese actors would likely find higher acceptance among those messaging apps' original user bases.

Emerging Use in Electoral Campaigns

Closed messaging apps have been widely used for political propaganda, social mobilization, and electoral campaigns in various countries worldwide since they started to become popular around 2014. The most popular messaging app worldwide, WhatsApp, is a fundamental communication vehicle for electoral campaigns in Latin America. Thus, the DFRLab aimed to see whether similar usage was emerging among Latino diasporas in the United States. However, since most of our analysis occurred between December 2021 and June 2022, outside of the electoral campaign season, we did not observe much activity in this regard.

As messaging app use continues to grow in the United States, we should expect their use for campaign activities also to rise. In Latin America, grassroots-level party organizing and get-out-the-vote operations are largely run using WhatsApp groups and distribution lists. Telegram channels are also used for campaign organizing in some countries. WhatsApp Communities, launched in November 2022, will allow greater reach, thus making messaging apps more effective for spreading electoral information. WhatsApp Communities can combine fifty groups for up to 5,000 members total for sharing announcements and conducting polls.[59]

Some Republican candidates have used Telegram for campaigning, while Democratic Party candidates were largely not using Telegram during our analysis period. This space may become important for the next presidential election, particularly in high Latino population states.



Screenshots of MAGA and Trumpist channels on Telegram posting on US elections. (Source, left to right: Dinesh D'Souza, May 24, 2022; Marjorie Taylor Greene (@realMarjorieGreene), September 3, 2022.)

59 "How to Create a Community," WhatsApp.

Unsolicited Sharing of Sexual Imagery and Content Derived from Child Abuse

A large part of the recent public conversation about messaging apps has been driven by concerns over their potential usage for harming children, particularly about the dissemination of child sexual abuse materials (CSAM).

Initially, this project did not intend to focus on this sort of harmful and illegal content because our focus was on overall usage. However, during our WhatsApp research, we came across CSAM, as well as sexual imagery of young adults.[60] We found some sexually explicit imagery, including CSAM, on English-language WhatsApp public groups that we joined through links posted on Reddit or Facebook. Those groups were promoted as groups for general conversation, business opportunities, or sharing entertainment content. Most of the sexual content found consisted of short videos of young women or girls (some potentially underage) performing sexual or sexualized activities. Some of these videos prompted viewers to privately message the posting account to supposedly contact the young women in the videos. We also found adult male homosexual content, but it was less common. During our analysis, we encountered—and subsequently reported[61]—a few instances of unmistakable CSAM. There were very few cases, but we encountered instances of sexual abuse of boys that appeared to be between five and twelve years old.

In the public groups we observed, participants typically did not respond or react to sexually explicit content. Occasionally, a few participants left the groups immediately after such content was shared, but we did not observe users confronting those who sent it. There is no way to know if these same users reported the objectionable content before leaving, nor is there any way to tell how many users, if any, reported objectionable content while remaining in the group.

Unsolicited Messages from Business Accounts

While organizational accounts, such as business or premium accounts, have some benefits—e.g., covering local news, delivering public service information, enabling grassroots organizing, growing small- and medium-size businesses—we also encountered more problematic behavior by some of the business accounts that we identified during the course of our research.

In particular, we received several unsolicited private messages from small-business accounts participating in the WhatsApp public groups that we were observing. We also received a few unsolicited cryptocurrency and multilevel marketing offers from unknown Telegram users after joining public groups focused on matters related to COVID-19.

Some businesses also joined public groups to gain access to participants in order to push their products. Without tougher platform policies against spam and data-protection standards on messaging platforms, this type of activity by businesses is likely to continue proliferating.

60 During this research, the DFRLab did not find explicit sexual imagery or images that may derive from sexual exploitation on WeChat or Telegram public channels.

61 When members of the DFRLab found CSAM while observing public groups on WhatsApp, they used the in-app user reporting feature to inform the platform of the unsolicited "offensive message." The menu does not include specific options such as CSAM, false information, hate messages, or suspicion of criminal activity.

Methods for Detecting Harmful Content in Messaging Apps

Messaging platforms use different techniques for monitoring usage and compliance with policies on acceptable content. Platforms' means for detecting harmful content in encrypted messaging apps can rely on three types of *content oblivious* methods: user reporting or flagging, analysis of metadata, and analysis of behavioral signals. These three methods are *content oblivious* in the sense that the platform does not access the content of user communications.[62]

Some messaging apps employ content scanning and flagging, which is *content dependent*, as is the case with WeChat, which deploys automated monitoring for all user conversations on the platform.[63] Automated content scanning and flagging is not possible on encrypted messaging apps.

User Reporting

The most straightforward way to handle abuse monitoring in encrypted messaging apps is user-initiated reporting. Almost all messaging platforms enable user reporting of spam or abuse, as was the case of the three platforms comprising this research. Most messaging apps provide in-app user reporting for spam and harassment. In the three messaging apps studied, users can report other sorts of content or activities that violate terms of service, including CSAM, grooming, unsolicited or nonconsensual sexual content, self-harm, hate speech, calls for terrorism, and disinformation. Platforms leverage user reporting to be able to review content shared that would otherwise be encrypted to them.

For several years, Telegram users have been able to block or report other users for spam, calls for violence, child abuse, and unsolicited pornography. By September 2022, Telegram added fake accounts (or impersonation), selling illegal drugs, and publishing personal details (doxing) to the reporting motives. When users press the "Report" button in a Telegram chat, they forward the selected message (or messages) to the app moderators. If the moderators find that the messages violated the terms of service, the infringing account becomes limited from contacting other users temporarily. Telegram is among the few messaging apps that offer an appeal channel for users sanctioned for spam or other violations through their Spambot chat.[64]

WeChat users can report other users for illegal activities. Reporting reasons include scams and fraud, obscene content, illicit sales and gambling, violence and terrorism, political rumors, and the vague reason of "compromised account." The user must select a report reason and provide the related chat scripts and images as evidence for submitting the report. Then the platform reviews the report and notifies the reporting user of the result via the official account WeChat Team.

As mentioned before, WhatsApp allows users to report accounts, groups, or messages that violate terms of service, including sharing prohibited content. People can report groups or users by selecting that option on their contact cards. Reporting a message only entails pressing and holding in a message to display a menu and selecting *Report*. According to WhatsApp's FAQ, the platform receives the reported group or the user ID, the last five messages sent by the reported user, the message date and time, and the type of message (image, video, text, audio).[65] User reports are sent to an automated queue for trust and safety personnel to review and to decide on sanctions if a policy was violated. As a result of these reports, WhatsApp may deactivate groups or ban users. The person reporting it will not, however, receive a response from the company about any measures taken.

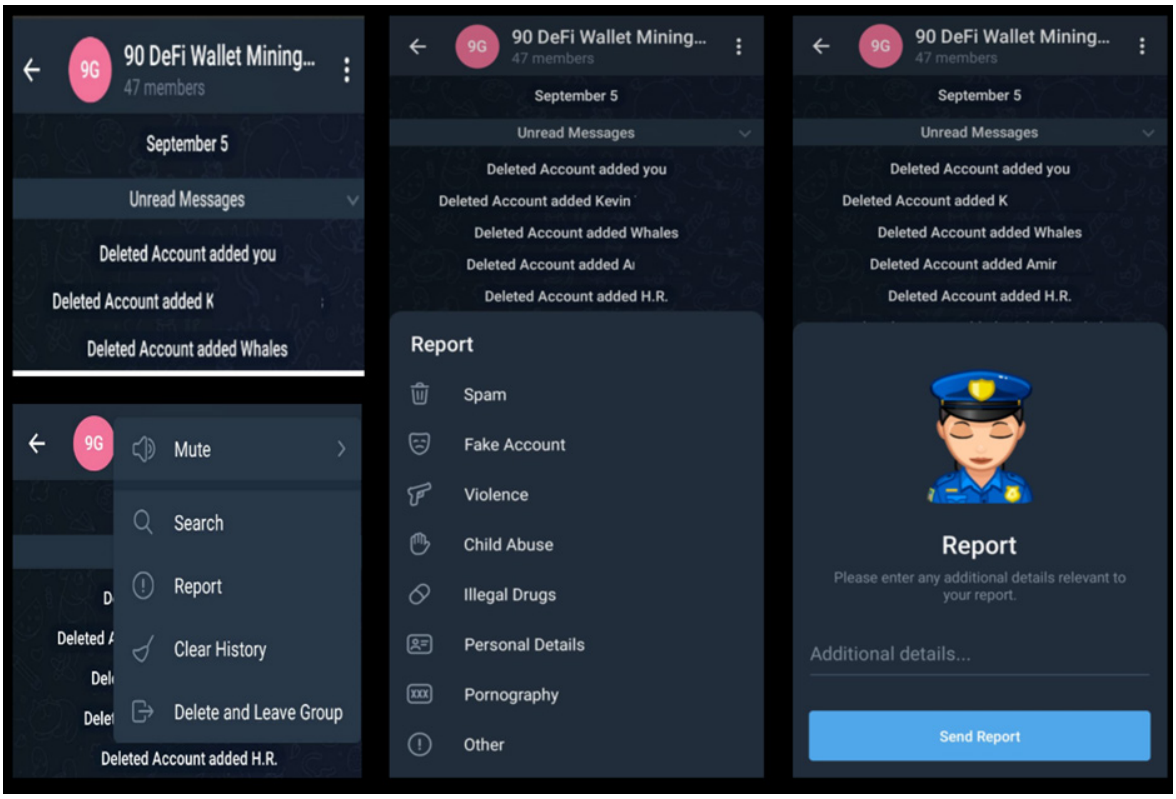
Besides reporting abuse to the messaging platform, users may report harmful or illegal content directly to law enforcement agencies. Users also can resend potential misinformation to fact-checking services. For example,

62 Riana Pfefferkorn, "Content-oblivious Trust and Safety Techniques: Results from a Survey of Online Service Providers," *Journal of Online Trust and Safety* 1.2 (2022).

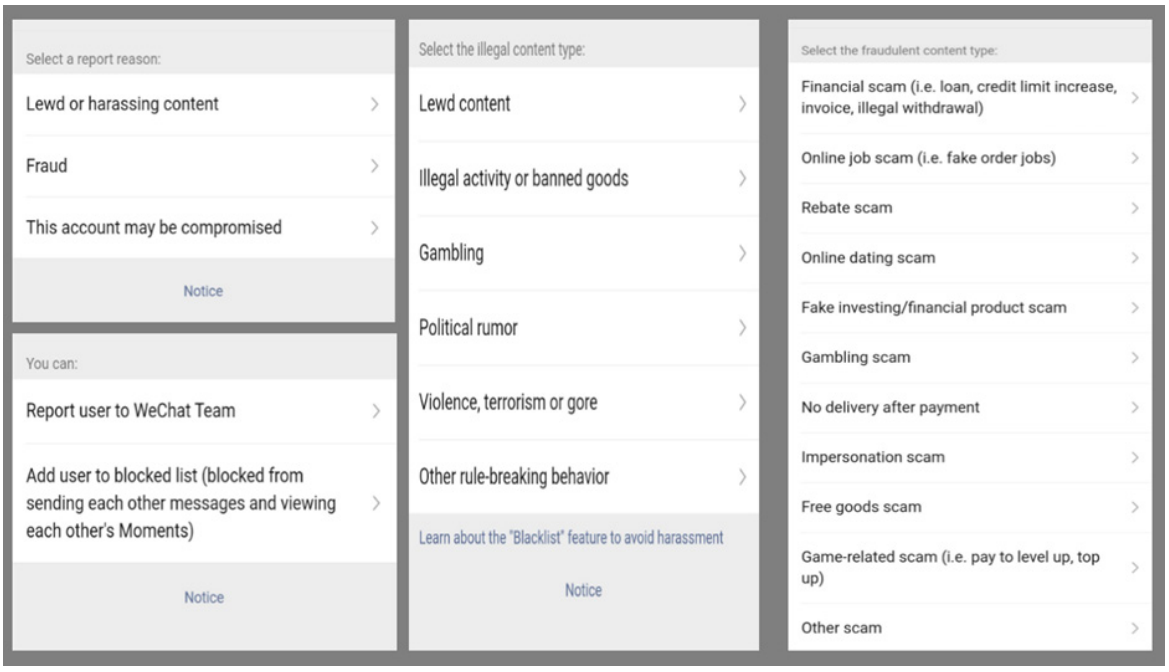
63 Jeffrey Knockel et al., *We Chat, They Watch: How International Users Unwittingly Build Up WeChat's Chinese Censorship Apparatus,* Citizen Lab Research Report no. 127, University of Toronto, May 2020, https://citizenlab.ca/2020/05/we-chat-they-watch/.

64 "Telegram Privacy Policy," Telegram (under 1.0), August 14, 2018, https://telegram.org/privacy#5-3-spam-and-abuse.

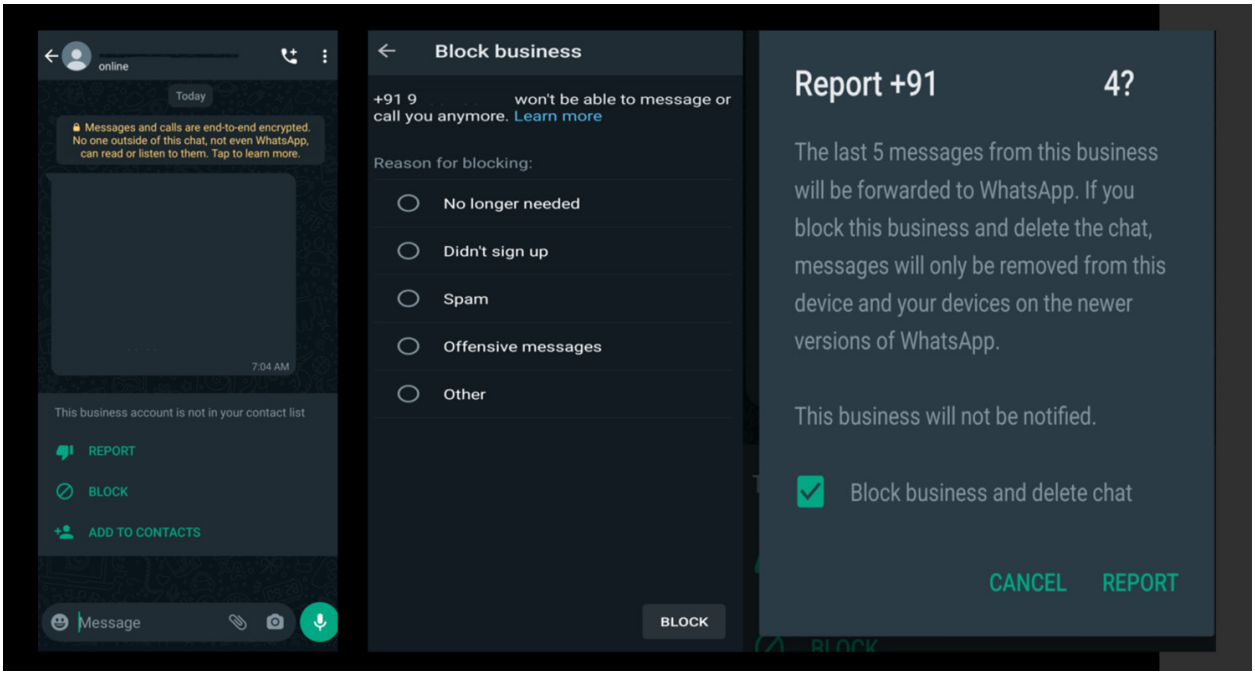
65 "About Blocking and Reporting Contacts," WhatsApp FAQ, n.d., https://faq.whatsapp.com/408155796838822/?helpref=faq_content.



Screenshots of the sequential screens that a Telegram user will see when choosing to block or report an undesirable contact. In this case, we were reporting an account that added a DFRLab monitoring account to a cryptocurrency chat without consent. The account had already been deleted when we reported, but the reporting process is the same. Reporting steps are not linkable. (Source: Telegram.)



Screenshots of the prompts a WeChat user will see on the English-language version of the app when choosing to report another user or a message. Reporting steps are not linkable. (Source: WeChat.)



Screenshots of the sequential screens that a WhatsApp user will see when choosing to block or report an undesirable contact. In this case, the DFRLab was blocking a business account that sent an unsolicited message to a monitoring account. A user reporting other regular users will see similar screens, though the reporting steps are not linkable. (Source: WhatsApp.)

for the 2022 US congressional elections, Meta partnered with Spanish-language media outlets Univision and Telemundo to launch fact-checking services on WhatsApp in which users could forward messages to automated chatbots to check their veracity. These user-generated reports of unacceptable content do not require breaking or "backdoor" access to the platform's encryption, since one *end* of the conversation is disclosing the message to the third party.

Message Franking

There are some developments from cryptography experts looking for technical solutions to addressing harmful content in encrypted spaces, including a technique called *message franking*.[66] Message franking is a solution for enabling cryptographically verifiable reporting of abusive or unacceptable content on encrypted messaging apps. Cryptographers ensure that message franking guarantees that a reporting user can prove to the platform to have received a given harmful or abusive message from another user. Simply put, message franking adds a tag to the message that the platform can decrypt to verify sender and recipient identifiers. Message franking assures that a user cannot claim having received a harmful or abusive message from another user if they in fact have not.[67] This technique guarantees message authenticity, the correct attribution of sender and receiver, and encryption integrity, since franking does not provide the encryption keys but rather authenticates the message information. Facebook Messenger, for instance, implemented message franking in 2017.[68]

Metadata and Behavioral Analysis

Several messaging platforms conduct analysis of metadata to detect what the platforms refer to as usage in violation of terms of service, ranging from spam to

66 Seny Kamara et al., *Outside Looking In: Approaches to Content Moderation in End-to-End Encrypted Systems*, Center for Democracy & Technology, 2021, https://cdt.org/insights/report-outside-looking-in-approaches-to-content-moderation-in-end-to-end-encrypted-systems/.

67 Kamara et al., *Outside Looking In*, Center for Democracy & Technology.

68 Facebook (now Meta), "Messenger Secret Conversations—Technical Whitepaper," Version 2.0, May, 2017, https://about.fb.com/wp-content/uploads/2016/07/messenger-secret-conversations-technical-whitepaper.pdf; for illustrations of how message franking works, see N. Tyagi et al., *Asymmetric Message Franking: Content Moderation for Metadata-Private End-to-End Encryption*, 2019, https://www.cs.cornell.edu/~tyagi/slides/amf.pdf.

criminal activities. Metadata include information about the origin of the data (device, user location), its structure or format, and how it was shared. Encryption protects the content exchanged but does not protect the metadata that is used to pass the content from sender to recipient. Both WhatsApp and WeChat acknowledge that their platforms run metadata analysis to enforce their acceptable content policies as well as to monitor app performance. Telegram states that the platform only uses a limited amount of user data to provide the service, guaranteeing security and spam mitigation.

Metadata analysis can be employed to identify accounts that may be likely to be spreading spam, malware, or CSAM. Metadata also can be practical way to identify accounts undertaking inauthentic behavior, particularly if it involves any automation of posting, sharing, or forwarding. Some metadata, such as communication logs and geolocation, may allow partial tracing of harmful content propagation. The most helpful metadata for detecting unacceptable content may be the types and frequency of account actions (e.g., messages sent, number of groups joined, media formats).

WhatsApp acknowledges applying machine learning techniques to metadata, group descriptions, and group icon photos to assess suspected CSAM. The platform also indicates that they have implemented text-based classifiers trained on CSAM to run against groups names and descriptions that can be flagged for human revision. After detection, offender groups can be immediately deactivated, and administrator accounts may be banned from using the platform.[69]

Meta has also disclosed that the company has detected "malicious coordinated behavior in end-to-end encrypted messaging by connecting actors' cross-platform and behavioral signals."[70] Behavioral signals refer here to actions that users perform on Facebook, such as joining groups, liking pages, visiting other user profiles, requesting friendship, or direct messaging. If Facebook detects assets engaging in coordinated inauthentic behavior on its main social media platform, and those assets have phone numbers tied to WhatsApp accounts, Meta can pass that information to the messaging platform. Then, WhatsApp can review associated metadata to determine if inauthentic behavior has also been coordinated in the messaging app. In cases where confirmatory evidence is gathered, WhatsApp might then ban those users.

In some cases, machine-learning procedures applied to metadata and behavioral signals may be helpful for detecting never-before-seen harmful content. For instance, by using machine learning, it is possible to flag users that join many public groups, send many pictures and videos, rarely reply to other users in the groups, and receive several blocks after sending media as potential CSAM spreaders deserving further investigation. Telegram and WeChat do not provide in their public documentation detailed information about how they use metadata analysis or other automated data analysis techniques; however, WeChat acknowledges running metadata analysis to ensure compliance with its terms of usage.[71]

Exceptional Access Backdoors or Key Escrow Systems

Law enforcement agencies around the world regularly request "exceptional access" to encrypted services. In the United States, such requests are usually made for investigations stated as affecting national security.[72] These demands are almost as old as the web, as law enforcement agencies in the United States have asked for "exceptional access" to encrypted services at least since the mid-1990s,[73] alleging different pressing issues over the years.

When it comes to messaging apps, some law enforcement and national security experts favor "key escrow," a backup decryption capability that allows authorized persons (e.g., government officials) to obtain a recovery key for decrypting ciphered content. Key escrow systems work as a master key that is kept in a vault but that can open the doors of each house in town following law enforcement requests. Proponents of such systems

69 "WhatsApp Help Center—How WhatsApp Helps Fight Child Exploitation," WhatsApp, 2021, https://faq.whatsapp.com/general/how-whatsapp-helps-fight-child-exploitation/?lang=en.

70 Business for Social Responsibility, *Human Rights Impact Assessment: Meta's Expansion of End-to-End Encryption*, 2022, https://www.bsr.org/reports/bsr-meta-human-rights-impact-assessment-e2ee-report.pdf.

71 "WeChat—Terms of Service," WeChat, March 22, 2022, https://www.wechat.com/en/service_terms.html.

72 Mieke Eoyang and Michael Garcia, "Weakened Encryption: The Threat to America's National Security," *Third Way Cyber Enforcement* (series), September 9, 2020, https://www.thirdway.org/report/weakened-encryption-the-threat-to-americas-national-security.

73 Wendy Grossman, *net.wars,* (New York: New York University Press, 1997).

argue that law enforcement and national security agencies should be allowed to obtain decryption keys for pursuing their investigations.[74] WeChat provides backdoors that allow law enforcement to access user data when needed, according to Chinese legislation.[75] In early 2023, suspicions about Russian Federal Security Service (FSB) access to Telegram's private and even secret chats resurfaced, though those suspicions remain unsubstantiated by evidence.[76]

Aside from such a model effectively negating user expectations of privacy when using messaging platforms, the problem with key escrow or other approaches is reliance on an assumption that such a key will never be obtained by bad actors and that any government with a key will use it only within strictly legal parameters. Both of these assumptions are hard to sustain, according to security analysis conducted by independent experts.[77]

Automated Scanning and Hash Databases

For years, content-moderation advocates have been promoting the implementation of automated detection of potentially harmful content within all sorts of internet-enabled exchanges, including messaging apps. In particular, groups seeking to fight CSAM and countering online terrorism have been advocating for the deployment of automated scanning techniques, either server or client based. As outlined below, however, such automated scanning would nullify the privacy protections afforded by E2E encryption.

Preemptive detection methods, such as server-side or client-side scanning, attempt to match content a user is sending against a hash database of previously identified potentially harmful or illegal content.[78] Matched content can, for example, be automatically blocked from upload or subjected to a sanction established by the service provider's policies. Client-side scanning could potentially report an instance of hash-matching directly to a law enforcement agency or a watchdog organization besides the service provider.

However, both server-side and client-side scanning are ineffective for identifying never-seen-before harmful or illegal content that is not already part of a database. Currently, hashes are available for terrorist and violent extremist content included in the Global Internet Forum to Counter Terrorism's database and CSAM in the National Center for Missing & Exploited Children's database.

The Global Internet Forum to Counter Terrorism (GIFCT), a nonprofit founded by Facebook, Microsoft, Twitter, and YouTube to explore technical solutions to counter terrorist and violent extremist activity online, hosts a shared database of identified terrorist content. The GIFCT hash-sharing database stores hashes of terrorist content detected on members' platforms. GIFCT has a taxonomy for database inclusion that considers whether the content producers are designated terrorist entities according to the United Nations or declared perpetrators of a terrorist incident according to the organization's content incident protocol. The hashes in the database are labeled as an imminent credible threat, graphic violence against defenseless people, glorification of terrorist acts, recruitment and instruction, and perpetrator content.[79]

The GIFCT database does not store personal identifiable information of any users associated with member platforms, only content hashes. Only member tech companies have access to this hash-sharing database. WhatsApp and Discord, as well as several social media platforms, are now GIFCT members; neither Telegram nor WeChat (or other messaging platforms) have joined. GIFCT does not share data with law enforcement agencies.

The National Center for Missing & Exploited Children (NCMEC) is a nonprofit organization established by the US Congress to serve as a clearinghouse and reporting center for all issues related to the sexual exploitation and abuse of children. NCMEC shares all reports it receives with law enforcement and aims to prioritize response to ongoing child abuse cases. The NCMEC maintains a hash database of CSAM. This database has a tip line that receives reports from the public, including survivors. However, 99.3 percent of the reported CSAM in 2021

74 Kamara et al., *Outside Looking In,* Center for Democracy & Technology.

75 WeChat, "Law Enforcement Data Request Guidelines," October 15, 2019, https://www.wechat.com/en/law_enforcement_data_request.html.

76 D. Loucaides, "The Kremlin Has Entered Your Telegram Chat," *Wired*, February 2, 2023, https://www.wired.com/story/the-kremlin-has-entered-the-chat/.

77 A. Stepanovich and M. Karanicolas, "Why an Encryption Backdoor for Just the 'Good Guys' Won't Work," *Just Security* (online forum), Reiss Center on Law and Security, New York University School of Law, March 2, 2018, https://www.justsecurity.org/53316/criminalize-security-criminals-secure/.

78 A hash is a unique, fixed-length string of random numbers and letters that is generated to identify a file.

79 GIFCT, *Broadening the GIFCT Hash-Sharing Database Taxonomy*, July 2021, https://gifct.org/wp-content/uploads/2021/07/GIFCT-TaxonomyReport-2021.pdf.

came from voluntary detection efforts by online electronic service providers.[80] These service providers run hash-value matching on their own servers. The biggest contributors to NCMEC hash database are Facebook, Instagram, WhatsApp, Snapchat, Google, TikTok, and Twitter. Facebook reported 76 percent of the CSAM added to the database in 2021, while WhatsApp reported 5 percent.[81] The other messaging apps that contribute to the database are Kik and Wickr; gaming platform Discord, which has a direct-messaging feature, also contributes.

WhatsApp indicates that the platform proactively removes accounts or groups that share CSAM when users report these accounts or when CSAM is revealed by scanning group icon photos.[82] Indeed, scanning public surfaces such as group icon pictures does not impinge on encryption.

Implications of Different Methods for Content Detection on Messaging Apps

When point-to-point messaging apps become headline news, it is often due to their use by criminal elements to publicize terrorist acts, spread inappropriate imagery, or other high-profile harms (which are also propagated across other internet platforms). As a result, the public discourse about messaging apps can be dominated by conversations around how to counter terrorism or prevent the spread of CSAM. Given the severity of these cases, it is no surprise that issue advocates and law enforcement officials turn quickly to calls for companies to more stringently police their platforms for such content.

These stakeholders often argue that messaging apps are spaces that allow criminal actors to hide their activities. Thus, they ask for privileged access or backdoors for law enforcement, tools for automated monitoring, or introducing human content moderation in messaging apps. Security experts and data protection advocates warn that these approaches introduce risks to the overarching security of these systems and can be misused by governments and nonstate actors alike. When it comes to E2E encrypted messaging apps, any third-party access introduces vulnerabilities into the system, increasing risks of unauthorized adversarial access to information that the parties in a conversation want to keep confidential. As the Global Encryption Coalition states, "There is no way to make a door that only the 'good guys' can open and the 'bad guys' cannot. Put differently, encrypted messaging with a backdoor for law enforcement is no longer encrypted messaging, as such apps become as insecure as text messaging via SMS. Creating a backdoor weakens the security of the whole system and puts all its users at risk."[83]

Others have proposed content-dependent preemptive techniques, such as server-side or client-side scanning . In this approach the server or client-side scanning is used to match content a user is sending against a database of previously identified potentially harmful content. But this too would necessitate third-party access to a user's device. Security experts warn that even client-side scanning that only reports positive or negative matching to a third party (e.g., the platform or law enforcement agencies) would compromise encryption integrity.[84] Methods for breaking encryption increase risks for all users, including children vulnerable to domestic abuse and sexually active teenagers engaged with other teenagers in romantic intimacy whose private pictures may be increasingly accessed by third parties seeking to use them illegally.

Additionally, hash-matching may be manipulated by adversarial agents to introduce false positives that compromise vulnerable users.[85] Moreover, attempts to trace content origin by message-hash may yield incorrect attribution since any minor modification in the message could generate a new hash value or code.[86]

Companies collecting and analyzing metadata do not break the fundamentals of encrypted communication, since the content remains between the two end users, i.e., the sender and the receiver. While this approach can illuminate valuable signals about potential abuse, it does not come without risk if not carefully protected with minimal access or sharing outside of threat teams. Usage

80 National Center for Missing & Exploited Children, "2021 CyberTipline Reports by Electronic Service Providers," 2022, https://www.missingkids.org/content/dam/missingkids/pdfs/2021-reports-by-esp.pdf.

81 Each Meta platform must report individually, as the parent company cannot report this type of data as a single entity.

82 WhatsApp, "How WhatsApp Helps Fight Child Exploitation," February 2021.

83 Global Encryption Coalition, *Breaking Encryption Myths—Global Encryption Coalition*, Internet Society and Center for Democracy & Technology, 2020, https://www.globalencryption.org/2020/11/breaking-encryption-myths/.

84 Alec Muffett, "A Civil Society Glossary and Primer for End-to-End Encryption Policy in 2022," 2022, https://alecmuffett.com/alecm/e2e-primer/.

85 Muffett, "A Civil Society Glossary and Primer."

86 Muffett, "A Civil Society Glossary and Primer."

and location data may convey compromising information about a user's social graph and visits to specific places. Data about who talked to whom, when, and for how long can be collected from messaging apps. Thus, metadata collection and analysis can also make people more vulnerable to surveillance and unauthorized disclosure of personal matters without proper controls.

In-app user reporting is the most privacy-respecting method for monitoring harmful content on messaging apps and perhaps the most effective. Nonetheless, there are technical challenges to guaranteeing report authenticity within encrypted apps and educating users about the importance of reporting.

Select Methods for Detecting Unacceptable Content in Messaging Apps

Method	Access to content	Encryption	Other implications
In-app user reporting	Content oblivious. The platform does not directly monitor user content.	Content can remain encrypted. An end-user reports it to the platform.	Users willingly engaged in conversations related to illegal activities are very unlikely to report harmful content to platforms.
Metadata analysis	Content oblivious. The platform does not directly monitor user content.	Content can remain encrypted. Platforms only analyze metadata.	False positives can be registered. Unacceptable behavior and content may go undetected. Sensitive personal data, such as geolocation, is gathered and analyzed.
Analysis of behavioral signals	Content oblivious. The platform does not directly monitor user content.	Content can remain encrypted. Platforms only analyze users' actions.	False positives can be registered. Unacceptable behavior and content may go undetected. Private, confidential, and legitimate interactions may be exposed. It may be repurposed for political surveillance.
Automated content scanning	Content dependent. The platform or other third party accesses the content.	Breaks E2E encryption model.	It weakens platform security making private communications more vulnerable to unauthorized access. It does not detect never-before-seen harmful content. Criminal actors can adapt content to avoid detection. False positives can be registered accidentally or because they are maliciously introduced by adversarial parties. Automated scanning software can be easily repurposed to monitor and censor legitimate content.
Backdoor access or key escrow systems	Content dependent. A third party accesses the content.	Breaks E2E encryption model.	It weakens platform security by making private communications more vulnerable to unauthorized access. Private, confidential, and legitimate interactions may be exposed. It may be easily repurposed for political surveillance.

Table showing the different methods unacceptable content can be detected by the platforms, including if the content is preemptively accessed, whether encryption is broken, and additional implications. (Source: Iria Puyosa, 2023.)

Key Takeaways

The intersection of technical features, policies around acceptable usage, and data gathering conducted by platforms leads to different message app models. There are underlying tensions between individual user preferences and governmental or societal demands for control. Most users may prefer a laissez-faire model until they are faced with unsolicited messages that they consider offensive or burdensome. Some democratic governments may want high-control models for addressing explicitly unlawful or terrorist activity, but authoritarian governments will use such models to repress dissent.

The messaging apps we reviewed may be similar in communication features but vary substantially in security, privacy, and content policies. E2E encryption by default in all types of interactions is the highest level of security offered by messaging apps: this is the case of WhatsApp. In contrast, nonencrypted messaging apps have the lowest level of security. Among the three messaging apps analyzed in-depth, WeChat has the lowest level of security, since this app only offers in-transit encryption.

Regarding data privacy, there are complex trade-offs that platforms must balance to protect their users and ensure app integrity. It is true that the less data a company has on its users, the lower the risk is of that data being accessed or misused by any number of actors. However, some data collection is essential to monitor app performance and safeguard users from abusive behavior by other users.

Similarly, most apps enact policies sanctioning harmful and illegal content. However, few messaging apps conduct extensive monitoring for unacceptable content, since human moderation and automated monitoring violates most apps' respective terms of service. Extensive automated monitoring and filtering is typical on WeChat, and the app acknowledges it in its terms of service. E2E encrypted messaging apps cannot monitor communication content to enforce their policies, since the platforms cannot decrypt content shared by their users. Thus, preventing the spread of harmful or illegal content on E2E messaging apps relies on user reporting.

Messaging app users use the apps to serve their needs and wants within the limits set by the technical features. Some users privilege security, confidentiality, and integrity, while others choose reach and adoption within their social networks. Moreover, users may use different apps for different purposes, depending on their needs. Most regular users tend to adopt the most popular messaging within their sociodemographic group, as we observed with Chinese diaspora using WeChat and Latinos using WhatsApp. Many high-risk and vulnerable individuals consider eavesdropping to be a threat and may take steps toward increasing security and privacy, such as using coded language, keeping sensitive personal identification details private, and avoiding the use of genuine photos and names in their messaging conversations. We found that migrants in WhatsApp public groups, for example, were adopting some of these tactics likely because of their vulnerability to victimization or exploitation.

Security, data protection, and privacy are more salient for high-risk individuals who are conscious of underlying threats (such as governmental surveillance) and their choice of app for communication. Human-rights defenders, abuse victims, and whistleblowers, for example, often seek E2E encryption since it adds a layer of security. Secure messaging mitigates threats of nonauthorized access to private conversations, adversarial actors' interference, and surveillance from governments, including foreign authoritarian governments engaged in transnational repression. Many high-risk individuals, like human rights defenders in repressive countries, depend on the security provided by encrypted messaging apps. Undermining encryption will place these individuals in life-threatening situations.

Our analysis of public groups and channels on WhatsApp, Telegram, and WeChat allowed us to identify a series of issues and topical trends in messaging apps related to public affairs. People are using public groups and channels to share news and discuss current events by exchanging content that they create, that they receive from contacts who do not participate in the groups, or that they reshare from social media and digital news outlets. Public groups and channels provide highly engaging spaces where users react quickly to the information they receive and can easily reshare to their networks.

Emphasis on the potential harms of the misuse of closed messaging apps has often ignored the benefits the same apps provide to their users. The DFRLab's analysis of public groups helped identify benefits that these

platforms provide to regular users in the United States. Among the most important are building and reinforcing community identity, enabling mutual support and resource exchange, and overcoming barriers to information. Groups also provide a space for diaspora communities to find shared support.

Conversations in encrypted messaging apps are better protected against surveillance, harassment, and hijacking than interactions occurring on public-facing social media platforms. Messaging apps are now major venues for discussing public affairs and political views since they offer features that facilitate sharing news, engaging in debates, and coordinating real-life activities. Since most of our analysis occurred outside of the electoral campaign season, we did not observe much activity regarding political campaigns and electoral mobilization (e.g., get-out-the-vote pushes), merely topical discourse, especially extremely active Trump supporters on Telegram anticipating the former president's campaign for the 2024 presidential election.

For politically motivated actors who aim to spread their messages (ranging from verifiable, fact-based information to disinformation), the main criteria for selecting a messaging app would be reach and amplification potential. Malign actors spreading disinformation and extremist content often seek the opposite of secrecy and privacy: extremist groups (e.g., the Islamic State group) seek attention, followers, and engagement, although some of them may use pseudonyms and other means to obfuscate their identity. Their goal is to propagate their beliefs and narratives, reinforcing the political identities of those who already sympathize with extremist ideologies but were previously silent and isolated.

The DFRLab observed that some features of Telegram public channels might exacerbate the spread of disinformation, such as large group sizes and lack of channel administrators' identity verification. On WeChat, mandatory identification and continuous automated monitoring may restrict both valid information and misinformation in public accounts to viewpoints and narratives aligned with the CCP. Meanwhile, the easy resharing of social media content and the enormous global user base enabled the spread of misinformation in WhatsApp. Still, smaller group sizes and group rules provide for countering manipulated content and restraining the circulation of misleading information on WhatsApp.

Closed messaging apps are already among the digital spaces instrumentalized for deploying foreign influence propaganda, similar to what happened with social media and state-affiliated media. The DFRLab found evidence of CCP influence over Chinese students on WeChat and of Russian narratives spread on Telegram. Still, we did not find evidence of systematic foreign influence campaigns on WhatsApp. However, since our observation sample was not representative, we could not rule out that such influence operations exist. Moreover, the risk of deployment of US-targeted political influence operations on WhatsApp will increase as the user base grows in the United States.

Although our research focused on channels whose members appear to reside largely in the United States and mainly discuss US domestic matters, we found that messaging apps enable transnational flows of information and opinions related to foreign or global issues. This happens partially through diaspora communities but also because ideological and identity communities transcend borders.

The growing presence of organizational or business accounts in messaging apps is a significant trend. Indeed, user demands have driven news distribution and business-to-customer interactions within messaging apps. Platforms have responded by formalizing these usage trends and providing premium features for organizational accounts. Nonetheless, organizational accounts pose new challenges regarding data protection and different risks of malicious use. During our research, we observed business users spreading harmful content, including misinformation and unsolicited sexual content.

Indeed, on WhatsApp, we found short videos of young women performing sexual or sexualized activities, prompting viewers to message privately to get in personal contact. We also found explicit male homosexual content shared in general conversation public groups. Upsettingly, we encountered some pieces of CSAM in public WhatsApp groups. CSAM and unsolicited sexual adult content are forbidden on the platform, and users can report the spreaders to have them banned from the app.

Messaging platforms employ different techniques for detecting harmful content or any practice that violates their acceptable usage policies. Closed messaging apps mostly rely on methods that do not require accessing users' communications such as user-initiated reporting or flagging, analysis of metadata, and analysis of behavioral signals. The most common and least intrusive way to detect abuse in messaging apps is in-app user reporting.

However, increasing in-app reporting effectiveness to tackle harmful content may require enhancing interfaces to make these tools more prevalent on apps' interfaces and improving digital literacy. Several messaging platforms, such as WhatsApp, also conduct analysis of metadata to detect unacceptable usage and apply machine-learning techniques to metadata and behavioral signals. These advanced techniques can be effective in detecting spam, malware, CSAM, and coordinated spread of disinformation. However, these techniques may be balanced to keep protecting user privacy and security.

Content-dependent preemptive methods, such as server-side or client-side scanning to match content a user is sending against a database will compromise encryption integrity, weaken security, and erode privacy protections. Furthermore, these techniques would be inefficient to detect never-before-seen content, which would continue to depend on user reporting.

Law enforcement demands present an explicit challenge to user security and privacy. Setting key escrow systems or backdoors for law enforcement "exceptional access" is damaging, as these systems break encryption and introduce serious vulnerabilities in personal communications. Besides, breaking encryption would not eradicate the circulation of harmful content, which will continue in unencrypted social media and on the dark web. Backdoors undermining encryption would, however, expose high-risk individuals to an increased likelihood of suffering harm by exposing their personal information.

Recommendations

Based on this research, the DFRLab identified a number of recommendations for both platforms and governments. These recommendations span product, policy, and regulatory interventions, centering user needs, agency, and security. As our research focused on the United States, so too do these recommendations. However, the transnational nature of these platforms and information flows in general mean that actions taken in the United States and in the design of platforms will inevitably be impacted by and affect internet-based communications worldwide.

Recommendations for Platforms

Since platform governance needs to take into consideration the challenges of competing demands, an assessment of trade-offs and the impacts on human rights should be an integral part of designing secure products and enforcing policies on acceptable usage. The DFRLab recommends the following for platforms to align their policies and product design as a means of reinforcing user security, privacy, and trust. The closed messaging app platforms should:

- **Invest in practical in-app reporting tools for harmful activities.** Product changes could have a significant impact. Platforms could make reporting menus accessible on the main messaging screen, and clear definitions of harmful or unacceptable content or behavior available in the app's help section. Platforms could also notify users about the course of action taken as a result of their reporting of unacceptable content or behavior and could add appeal procedures to their policies. Other helpful steps might include updating reporting categories to reflect the most commonly detected harms with the most serious harms, such as CSAM, displayed at the beginning of the list of reasons to report. Platforms could also send prompts reminding users in large public groups of unacceptable content and available reporting tools. As a best practice, the ecosystem would be helped by platforms including detailed statistics of in-app user reporting in their transparency reports.

- **Invest in testing the effectiveness of giving group administrators moderation privileges.** In most apps, administrators can set group guidelines on topics and acceptable manners and ban users, as happens on the apps that we assessed in this report. Platforms should run a pilot test assigning additional moderating privileges to group administrators: e.g., the ability to flag messages that violate group guidelines, remove objectionable content, mark messages as not allowed for forwarding, and mute a participant for a defined amount of time. Data from pilot testing of these interventions should be made available to researchers who could assess their effectiveness.

- **Optimize data collection and processing to maintain service integrity, enhance user safety, and guarantee privacy.** Metadata collection and analysis are necessary for user safety within messaging apps, particularly in public groups. Transparency notices explaining data collection and processing must be readily available to users. Metadata should be stored on secure encrypted servers for the time required to provide services, safety analysis, and vetted research. Special safeguards should be provided regarding the most sensitive data, such as geolocation. Messaging platforms may retain data from accounts that have triggered a CSAM or terrorism report for extended periods than typically established in their data retention policies when criminal investigations are pending. For messaging apps that offer E2E encryption, platforms should guarantee that servers are securely encrypted and open their encryption protocols to independent audits.

- **Collaborate with counterterrorism initiatives.** Messaging apps, including those that are E2E encrypted, should join industry initiatives such as the Global Internet Forum to Counter Terrorism (GIFCT) and Tech Against Terrorism. Currently, WhatsApp is the only E2E encrypted messaging app that has joined GIFCT. By joining GIFCT, E2E encrypted messaging apps could share technological approaches to detecting terrorist activity and collaborate in addressing risks and needs for responding to terrorism incidents. As GIFCT members, messaging apps will have access to the hash-sharing database. Messaging platforms should contribute with hashes of terrorist images found on unencrypted surfaces, such as profile and group photos, and content included in user reports.

- **Define robust policies for business and organizational accounts, including media outlets, political organizations, nonprofit organizations, and government entities.** Although organizational accounts are still a minority in messaging apps, their use will grow exponentially in the future, driven by user demands and new features incorporated into the platforms. Messaging platforms should start preparing their trust and safety workflows for the issues this type of account brings, including the risks of massive data breaches from in-app transactions. As business or organizational accounts become more prevalent, platforms should remind users of the differentiated privacy and data breach risks entailed in interacting with these accounts. Organizational accounts should provide documentation that certifies what type of organizations they are and which services they aim to deliver to clients via messaging apps. Moreover, they should disclose whether they will delegate their messaging operations to a vendor and the level of access to user data they will provide to other organizations in their supply chains. Organizational accounts should seek consent for retaining user data and for any metadata analysis they plan to conduct. Also, they should ensure security measures are in place to protect messaging interaction data from unauthorized access.

- **Allow users to customize their privacy settings for different types of interactions.** Messaging app users should be able to customize and adjust their privacy settings at different levels when interacting with another individual, a group, or a business. Users should be able to set trusted contacts lists or block some contacts from seeing status. Privacy features should be deployed, such as marking certain messages as not allowed to be forwarded or preventing chat screenshots, in order to diminish the unauthorized resharing of sensitive content.

- **Partner with outside researchers and investigation centers.** Currently, US legislators are considering various bills that would require platforms to share data with researchers, notably the Platform Accountability and Transparency Act (PATA) and the Digital Services Oversight and Safety Act. Messaging platforms should proactively advance protocols for data sharing that go beyond such transparency mandates. Sound, independent research should provide findings and insights for understanding user behavior, emerging harms, and potential solutions. Research priorities should be defined jointly with relevant stakeholders, with researchers specifying what type of data can be most helpful to understand, explain, or predict critical phenomena, such as the spread of disinformation or the instigation of political violence. For their part, the platforms should ensure quality, granularity, and data security. As data stewards, platforms should develop protocols for sharing aggregated metadata with vetted outside researchers and investigation centers studying messaging usage. Messaging platforms can build virtual labs where researchers can analyze datasets without downloading user data to insecure servers. In any case, datasets should not include personal identification data, including a user's precise geolocation.

- **Consider human rights impacts when designing content and usage policies.** International standards of freedoms and rights may guide the drafting and revision of messaging app policies. The United Nations Guiding Principles on Business and Human Rights provides a high-level framework for principles applicable to messaging app operations regarding the right to privacy and freedom of association and assembly. Congruently, platforms should conduct human rights impact assessments before and after rolling out significant new features or enacting new policies to avoid unintentionally harmful effects. Using ethical design checklists alongside product design processes could help foresee potential harmful effects that would require policy interventions later if left unaddressed.

- **Address trust and safety holistically.** Messaging apps' integrity, trustworthiness, and security require balancing conflicting demands. Centering user choice when designing features and user rights when defining policies could help platforms to handle these tensions. Trust and safety high-level guidelines should assess whether a solution creates or exacerbates another problem. Also, platforms should make adjustments during product testing and implementation to mitigate foreseeable harms.

Recommendations for Policymakers

The perennial challenge for lawmakers is to bridge siloed policy conversations across interlocking jurisdictions and committees. As this report shows, policies intended to address seemingly targeted issues such as terrorism could unintentionally undermine key elements of the internet itself, such as calls to gut encryption. Thus, the DFRLab recommends policymakers address emerging issues related to the usage of messaging apps in a holistic manner—viewing the internet as the interlocking digital

ecosystem it is. This means advancing policies not only designed to protect individuals' rights, but also taking into consideration the potential international impact of regulations enacted in the United States.

- **Advance data protection legislation that sets rules for data collection, processing, storing, and sharing.** Data protection and privacy legislation are fundamental for addressing many of the policy issues arising in the messaging app ecosystem. The United States and other countries that do not have a federal data protection law face a more challenging situation when deciding where to draw lines between protecting user communication and enforcing policies on acceptable content and organizational usage of messaging apps. A successful data protection law would provide a general framework for more specific regulations in data-driven industries, including messaging apps or social media. The main aspects to be considered in data protection legislation include specific data collection purposes, data minimization rules, limited collection of sensitive data, limited storing of data according to the legitimate purpose, prompt deletion of geolocation data after the transaction requiring it ends, ensuring appropriate security for storing personal data and metadata, and the prohibition of sharing data with third parties without consent. Careful consideration must be given to which authority or agency, new or preexisting, should oversee any enforcement of such legislation, as well as issues related to state law preemption. In developing and passing such legislation, policymakers must consider the transnational flow of data and the extraterritorial scope of affected platforms' operations.

- **Avoid regulations that undermine encryption.** Regulations should not compel messaging platforms to conduct automated content scanning that breaks encryption, as proposed in a bill dubbed EARN IT (Eliminating Abusive and Rampant Neglect of Interactive Technologies Act of 2022). Platforms cannot be liable for content shared by users in encrypted messages, as the platforms do not have access to those messages by design. Escrow systems or the provision of backdoors to law enforcement and national security agencies attempting to access personal communications should be prohibited. Platforms should only provide access to a user's basic data and communications metadata when presented with a court warrant. Bulk requests for law enforcement access to user data and metadata should be prohibited.

- **Examine business practices and commercial services offered via messaging apps to identify regulatory gaps.** Regulatory bodies, such as the Federal Trade Commission (FTC), should assess how existing regulations cover business practices now being conducted using organizational accounts on messaging apps and whether regulatory gaps exist. Relevant existing legislation includes the Electronic Communications Privacy Act, the Stored Communications Act, and the Communications Decency Act. Any new regulations on messaging apps should prioritize protecting user data and acknowledge privacy expectations when interacting on closed or private chats. Policies and regulations affecting messaging app operations should incorporate consumer education and ensure that platforms provide users with accessible ways to report illegal or unacceptable activities.

- **Optimize resources for investigating criminal activities adjusted to the conditions of digital platforms.** Law enforcement should stay up-to-date on techniques for investigating criminal activities involving messaging apps and other digital services, including how to analyze metadata that platforms may provide after receiving a warrant. Following up on investigations tipped from hash databases (such as those maintained by the NCMEC) will be more respectful of law-abiding users' right to privacy and more effective than attempting to monitor every exchange in messaging apps. Law enforcement agencies should also submit CSAM seized during criminal investigations to NCMEC, contributing to enlarging the database with images not yet detected by platforms. Moreover, law enforcement must prioritize investigations of new content where children may be in physical danger and therefore focus on tracking down and arresting individuals exploiting and abusing children offline.

- **Promote digital literacy tailored to the risks faced by users of messaging apps.** Middle and high school curriculums should be developed to include digital literacy on messaging apps. Gamification could be helpful to convey age-appropriate content on harms that a child or teenager may encounter in messaging apps. Educational content should include digital safety when sharing in messaging groups, the risks of sexting, setting personal limits when interacting on messaging apps, and how to report abuses.

Conclusion

This report highlights how user choices on messaging apps interact with platform features and policies. We observed differences in messaging app use that may relate to app architecture, policies, monitoring, and enforcement by the platforms. The three apps we examined for this report represented a breadth of security and privacy attributes, from WhatsApp's E2E encryption by default to WeChat's automated monitoring of all content by default. Understanding platforms' respective acceptable use policies and what platforms promise regarding security, data protection, and privacy is a fundamental piece of any honest discussion concerning harms and risks within this environment.

In this report, the DFRLab presented a snapshot of a specific subset of messaging app usage in the United States. This investigation focused on public groups in WhatsApp, Telegram, and WeChat, but different dynamics and issues may be present in other messaging apps, encrypted or nonencrypted. Thus, expanding research to other apps with growing user bases is necessary.

Further investigation is still required on techniques for detection of harmful content and mitigation of abusive behavior on messaging apps. Emphasis should be placed on techniques for increasing the effectiveness of in-app reporting and metadata analysis.

The DFRLab initially anticipated to find more political content across the three messaging apps under observation, but it was not nearly as prevalent as anticipated. On Telegram, we did find conservative and far-right groups spreading ideological narratives, but we did not find US politics to be the subject of much discussion on WeChat or WhatsApp. Considering how messaging is used for political campaigns in countries with high adoption of these apps, similar usage may rise in the United States as the overall user base increases. Thus, a reexamination of messaging apps for public-facing political content during a period of higher political activity, such as ahead of the US 2024 presidential election, might yield better insight into their use for political discussions.

There are new threats and risks emerging from organizational account activities within messaging apps, including spam and scams, and the exposure of sensitive data. Early research on this topic could help platforms prepare their trust and safety workflows to deal with the issues organizational accounts bring with them, including issues affecting platform integrity and spreading harmful content.

Atlantic Council

Board of Directors

CHAIRMAN
*John F.W. Rogers

EXECUTIVE CHAIRMAN EMERITUS
*James L. Jones

PRESIDENT AND CEO
*Frederick Kempe

EXECUTIVE VICE CHAIRS
*Adrienne Arsht
*Stephen J. Hadley

VICE CHAIRS
*Robert J. Abernethy
*Alexander V. Mirtchev

TREASURER
*George Lund

DIRECTORS
Todd Achilles
Timothy D. Adams
*Michael Andersson
David D. Aufhauser
Barbara Barrett
Colleen Bell
Stephen Biegun
Linden P. Blue
Adam Boehler
John Bonsell
Philip M. Breedlove
Richard R. Burt
*Teresa Carlson
*James E. Cartwright
John E. Chapoton
Ahmed Charai
Melanie Chen
Michael Chertoff
*George Chopivsky
Wesley K. Clark
*Helima Croft
*Ankit N. Desai
Dario Deste
Lawrence Di Rita
*Paula J. Dobriansky
Joseph F. Dunford, Jr.
Richard Edelman
Thomas J. Egan, Jr.
Stuart E. Eizenstat
Mark T. Esper
*Michael Fisch
Alan H. Fleischmann
Jendayi E. Frazer
Meg Gentle
Thomas H. Glocer
John B. Goodman
*Sherri W. Goodman
Jarosław Grzesiak
Murathan Günal
Michael V. Hayden
Tim Holt
*Karl V. Hopkins
Kay Bailey Hutchison
Ian Ihnatowycz
Mark Isakowitz
Wolfgang F. Ischinger
Deborah Lee James
*Joia M. Johnson
*Safi Kalo
Andre Kelleners
Brian L. Kelly
Henry A. Kissinger
John E. Klein
*C. Jeffrey Knittel
Joseph Konzelmann
Franklin D. Kramer
Laura Lane
Almar Latour
Yann Le Pallec
Jan M. Lodal
Douglas Lute
Jane Holl Lute
William J. Lynn
Mark Machin
Marco Margheri
Michael Margolis
Chris Marlin
William Marron
Christian Marrone
Gerardo Mato
Erin McGrain
John M. McHugh
*Judith A. Miller
Dariusz Mioduski
Michael J. Morell
*Richard Morningstar
Georgette Mosbacher
Majida Mourad
Virginia A. Mulberger
Mary Claire Murphy
Edward J. Newberry
Franco Nuschese
Joseph S. Nye
Ahmet M. Ören
Sally A. Painter
Ana I. Palacio
*Kostas Pantazopoulos
Alan Pellegrini
David H. Petraeus
*Lisa Pollina
Daniel B. Poneman
*Dina H. Powell McCormick
Michael Punke
Ashraf Qazi
Thomas J. Ridge
Gary Rieschel
Michael J. Rogers
Charles O. Rossotti
Harry Sachinis
C. Michael Scaparrotti
Ivan A. Schlager
Rajiv Shah
Gregg Sherrill
Jeff Shockey
Ali Jehangir Siddiqui
Kris Singh
Walter Slocombe
Christopher Smith
Clifford M. Sobel
James G. Stavridis
Michael S. Steele
Richard J.A. Steele
Mary Streett
*Gil Tenzer
*Frances F. Townsend
Clyde C. Tuggle
Melanne Verveer
Charles F. Wald
Michael F. Walsh
Ronald Weiser
*Al Williams
Maciej Witucki
Neal S. Wolin
*Jenny Wood
Guang Yang
Mary C. Yates
Dov S. Zakheim

HONORARY DIRECTORS
James A. Baker, III
Robert M. Gates
James N. Mattis
Michael G. Mullen
Leon E. Panetta
William J. Perry
Condoleezza Rice
Horst Teltschik
William H. Webster

The Atlantic Council is a nonpartisan organization that promotes constructive US leadership and engagement in international affairs based on the central role of the Atlantic community in meeting today's global challenges.

1030 15th Street, NW, 12th Floor,
Washington, DC 20005
(202) 778-4952
www.AtlanticCouncil.org