Covert Campaigns:

Safeguarding Encrypted Messaging Platforms from Voter Manipulation

MARIANA OLAIZOLA ROSENBLAT, INGA K. TRAUTHIG AND SAMUEL C. WOOLLEY

NYU STERN
Center for Business and Human Rights

The University of Texas at Austin
Center for Media Engagement
Moody College of Communication

October 2024

Contents

Authors

Mariana Olaizola Rosenblat is a policy advisor on technology and law at the NYU Stern Center for Business and Human Rights.

Inga K. Trauthig is the head of research of the Propaganda Research Lab at the Center for Media Engagement (CME) at UT Austin.

Samuel C. Woolley is the Dietrich Endowed Chair of Disinformation Studies at the University of Pittsburgh and founder of the Propaganda Research Lab at the CME at UT Austin.

Paul M. Barrett, the NYU Stern Center's deputy director, provided assistance with the conception and editing of this report.

Executive Summary

Messaging platforms, which market themselves as spaces for private conversation, increasingly serve as arenas for intense political activity, including electoral campaigns. The architectures and features of applications like WhatsApp, Telegram, and Viber make them particularly useful as vectors for political propaganda, or information calculated to manipulate public opinion.

“The architectures and features of applications like WhatsApp, Telegram, and Viber make them particularly useful as vectors for political propaganda, or information calculated to manipulate public opinion.”

Although the presence of political content on these platforms has been known for some time, the mechanics of abuse, as well as effective strategies to counteract them, are not yet well understood.

Drawing on interviews with purveyors of political propaganda in 17 countries, a survey of over 4,500 messaging app users in nine countries, and insights from relevant experts, this report describes *how* political propagandists are exploiting specific in-app features to manipulate voters during elections. Lessons extracted from places where elections have just taken place (e.g., India, Mexico, and the European Union) can help inform mitigation strategies in the United States and other places where elections have yet to occur.

Empirical research on messaging applications is scant, largely due to the difficulty of studying platforms where some or all of the content is protected with end-to-end encryption, a cryptographic method that renders messages indecipherable for everyone except for the senders and intended recipients. End-to-end encryption offers propagandists useful concealment from electoral authorities and resilience to platform moderation. But the critical value of robust encryption to activists and dissenters at risk of surveillance by repressive regimes precludes eliminating or weakening encryption as an antidote to such propaganda efforts.

Given these competing stakes, this report suggests a pragmatic way forward. It offers a series of concrete recommendations for messaging platforms, policymakers, and researchers to help mitigate political manipulation efforts on messaging platforms without undermining the privacy guarantees of end-to-end encryption. The following is a summary of these recommendations.

Recommendations In Brief

To messaging platforms

1 **Establish strict account-creation limits.** To hinder the operation of fake accounts and "phone farms," platforms should impose restrictions on the number and pace of account creation. Where such restrictions already exist, platforms should address technical loopholes exploited by propagandists.

2 **Restrict large-scale broadcasting to verified channels and vetted business accounts**. Where a platform enables channels and business accounts to reach larger audiences, they should rigorously vet those accounts to ensure their authenticity and compliance with platform policies.

3 **Strengthen cross-industry and multi-stakeholder cooperation to identify inauthentic activity.** Messaging services should engage with peer platforms to identify cross-platform inauthentic activity and join cross-industry and multi-stakeholder initiatives.

4 **Support and improve access to accredited tiplines.** Platforms should inform users about the operation of accredited election misinformation tiplines, which provide independent fact-checking services to users on their apps, and enhance access to them through improved user interface design.

5 **Invest in user-driven fact-checking tools.** Platforms should develop tools for in-app fact-checking, such as one-click reverse image searches, and support research into other potentially effective tools.

6 **Choose whether to prioritize privacy or abuse mitigation—and be transparent with users.** Where privacy and abuse mitigation interests compete in the selection of app features and moderation approaches, platforms should make clear choices and be transparent with users about their implications.

7 **Bifurcate the platform's messaging service from its social media or broadcasting functions.** The private messaging service, protected with end-to-end encryption, should be reserved for individual and small-group chats. The social networking and broadcasting functions should be moderated rigorously, in line with mainstream social media standards.

To policymakers

8 **Do not impose legal obligations that undermine encryption.** Lawmakers should refrain from passing laws that require platforms to scan encrypted content for illegal material, trace forwarded messages to the original sender, or impose any other obligations that undermine the privacy of end-to-end encrypted communication. At the same time, lawmakers should require that messaging platforms provide some transparency as to their operation, policy enforcement, and business models—none of which compromises encryption.

9 **Support bottom-up media literacy efforts.** Policymakers should fund organizations that work collaboratively with communities to deploy context-sensitive media literacy strategies.

To researchers

10 **Contribute to media literacy initiatives.** News and civil society organizations should establish or support fact-checking and voter education tiplines.

11 **Develop ethical methodologies for studying encrypted online spaces.** Researchers should devise ethical methods for studying encrypted platforms with a view to enhancing the public's understanding of their effects on society.

1. Introduction

Arriving by private jet in Paris on August 24, 2024, Pavel Durov, the billionaire founder of the social platform Telegram, was greeted by police officers. The French government subsequently charged him with a long list of crimes related to the technology platform's failure to address illegal activity on the site, including enabling organized crime, child sexual abuse material, and fraud.[1]

> **"In many parts of the world encrypted messaging platforms have surpassed social media (i.e., Facebook, X, and TikTok) as the most widely used app category."**

The case, which was unfolding at the time of this writing, draws attention to the potential culpability of technology firms and their owners for illegal uses of their digital tools. It also highlights the dilemmas surrounding encrypted communication.[2] Durov's arrest has sparked outrage among technology advocates and free speech absolutists,[3] while drawing praise from law enforcement and children's safety groups. Many see his platform as critical to both privacy and free speech; others view it as a haven for predators and other malefactors.[4] It is not possible now to assess the legitimacy of the French government's criminal prosecution without knowing more about Durov's conduct. Was he directly involved in the spread of illegal content or the thwarting of legitimate government investigations, or is he being prosecuted because he passively allowed illegal content to circulate on Telegram? Either way, the case could have profound implications for technology firms and their regulation.

This report addresses another critical question: How can we safeguard encrypted messaging platforms from voter manipulation? Beyond illuminating the problem, the report offers a range of solutions, none of which involves compromising privacy of communications or undermining encryption. The report also relates the issue of voter manipulation to other harmful uses, such as extremist mobilization and fraud.

The broader context for this report is that messaging platforms like Telegram, WhatsApp, and Viber provide society with both clear benefits and challenges. They offer users an easy-to-use, often free, means of connecting with people around the globe. They allow for private, secure communication in repressive contexts. However, and often because they are secure, these platforms serve as channels for sharing harmful, and sometimes illegal, content. In France, Telegram enabled coordination and propaganda dissemination among the terrorists who committed the 2015 Paris attacks.[5] The use of messaging apps also affects democracy and democratic processes. Telegram has helped pro-democracy activists mobilize for protests in places like Hong Kong and Belarus.

Encrypted Messaging as a Sword and a Shield: Venezuela Case Study

While this report focuses on the use of encrypted platforms to manipulate voters, readers should keep in mind the value of encryption in enabling a secure means of communication for dissidents and human rights defenders in danger of state repression. Venezuela's July 2024 election, which was widely considered fraudulent and invalid, offers a useful case study on the benefits and limitations of encrypted communications technology in autocratic contexts.

WhatsApp, the most popular messaging app among Venezuelans, has long served as a key channel for information sharing and political mobilization, both within and outside of the country.[1] On election day 2024, WhatsApp groups disseminated citizens' reports and reactions to the high voter turnout alongside expressions of trepidation about the regime's potential interference with the results.[2] When the Maduro regime declared victory despite exit polls reportedly showing an opposition landslide, many community organizers used WhatsApp to plan and coordinate protests. Social media platforms including X also served as key distribution networks for information about those mobilization efforts.[3]

The regime was quick to respond. In the days following the election, there were multiple reports of security forces stopping citizens to check the content of their WhatsApp conversations and detaining those whose pictures or discussions revealed anti-government sentiment.[4] Concurrently, the regime encouraged supporters to delete WhatsApp from their phones and migrate to Telegram and Chinese-owned WeChat, seen as friendlier (and less privacy-protective) alternatives.[5] The regime also promoted a native non-encrypted app originally used for social service delivery, VenApp, to report on dissenting activities and intimidate demonstrators, going so far as to create an in-app tool called "Denuncia" (Spanish for denounce) designed for regime loyalists to report on their neighbors.[6] To throttle any remaining attempts at popular mobilization, the regime then moved to block access to Signal, which is favored by security-conscious journalists and dissidents, and X, a central platform for public-facing expressions of dissent.[7]

1 Puyosa, I., Azpúrua, A. & Suárez, D. (2024). Venezuela: A Playbook for Digital Repression. Atlantic Council. https://www.atlanticcouncil.org/wp-content/uploads/2024/07/Venezuela-a-playbook-for-digital-repression.pdf

2 https://proboxve.org/en/publicacion/keys-of-the-presidential-elections-in-venezuela-whats-going-on/

3 Interviews with Iria Puyosa, Roberta Braga and Mariví Marín.

4 https://freedomhouse.org/article/open-letter-technology-enabled-political-violence-venezuela

5 https://www.accessnow.org/press-release/maduro-keepiton-during-protest-and-unrest/

6 https://proboxve.org/en/publicacion/digital-terror-maduros-approach-to-silencing-critical-voices/; https://proboxve.org/en/publicacion/digitalterror-operation-knock-knock-on-the-hunt-for-opposition-voices/

7 Interviews with Iria Puyosa and Mariví Marín. See also https://www.theverge.com/2024/8/9/24217008/signal-blocked-venezuela-russia.

But it was also leveraged by the right-wing extremists who stormed the U.S. Capitol on January 6, 2021. Similar applications have been used in coordinated attempts to manipulate voters around the world. These are the issues we address here.

A prized political tool

Stepping into the Indian National Congress party's elections "war room" in New Delhi, one has the impression of entering a beehive. The room is abuzz with both volunteers and workers hired by the biggest opposition party in India to inundate voters on their phones with pro-party messages. They face a steep uphill battle. The Congress workers manage numerous computers and cellphones on each desk in a struggle to keep up with the ruling Bharatiya Janata Party's (BJP) massive online communications machine, dubbed the BJP Digital Army because of its dedicated and coordinated cadre.[6] The key battlefield is WhatsApp.

WhatsApp, an instant messaging app, has served as one of the most important electoral weapons in India since 2014, when incumbent Prime Minister Narendra Modi made it central to his political campaign.[7] While cementing itself as a crucial mode of communication in the country (it is the most popular messaging platform in India by far[8]), the app became a prized political tool.

In its original form, WhatsApp was a basic messaging service with a simple user interface that enabled one-to-one text messaging.[9] After its acquisition by Meta (then Facebook) in 2014, the initial application expanded rapidly, both in terms of its user base and features. Today, WhatsApp can be characterized as part-messaging, part-social media platform. It has features ranging from individual chats, to "communities" comprising up to 100 group chats, to "channels" for broadcasting content to large audiences. The platform provides an inexpensive and effortless mode of communication—in some countries, such as Brazil, WhatsApp usage does not count against consumers' pre-paid

data allowance and is therefore free.[10] Correspondingly, the app allows politicians and their operatives to reach voters directly and en masse, merging a sense of intimacy with the potential for virality.

Modi and his BJP are not alone in grasping the potential of WhatsApp for political propaganda and persuasion.[11] Nor is WhatsApp the only messaging app that can be exploited in these ways. Ukrainian officials use Viber to spread campaign messages to small groups of voters and Telegram for larger information operations. U.S. political strategists also target diaspora communities across various states via Telegram and WhatsApp.[12]

Many messaging apps, including WhatsApp, have additional value to political operatives because they apply end-to-end encryption across most or some part of their platform. End-to-end encryption is a cryptographic method that allows only the "ends" of a communication—the senders and intended recipients—to access the content of messages.

Selected messaging platforms: Encryption, Features, and Business Models

	Level of encryption	Relevant features	Business model
WhatsApp	**Partial** End-to-end encryption (E2EE) applied by default to all features except channels; metadata is also not E2EE.	• Individual chats • Group chats (max 1,024 members) • Communities (i.e., a "supergroup" of max 100 groups) • Intra-app forwarding • Status updates • Broadcast lists (up to 256 individual contacts) • Channels (unlimited audience; not E2EE) • WhatsApp Business API • Generative AI-powered chatbot	Shares user metadata (e.g., phone number, profile photo, status update, IP address) with Meta for ad targeting. Additional revenue from Whatsapp Business API and Whatsapp Pay.
Telegram	**Low** E2EE only for individual "secret chats;" all other app functions are not E2EE.	• Group chats (max 200,000 members) • Intra-app forwarding • Channels (unlimited audience) • Secret chats (one-to-one chats under E2EE) • Global search function allowing users to search groups • Newsfeed • Stories • Chatbots • Telegram for Business	Original funds came from founder's assets; some revenue from premium subscriptions and ads in public channels.
Signal	**Full** E2EE applied to all content and metadata.	• Individual chats • Group chats (max 1,000 members) • Intra-app forwarding • Stories • Stickers	Nonprofit run by Signal Foundation; initial donation by former head of WhatsApp; small donations from individual donors.
Viber	**Partial** E2EE applies only to individual chats, group chats, and 1-on-1 calls; all other features have encryption-in-transit.	• Individual chats • Group chats (max 250 members) • Intra-app forwarding • Communities (groups with unlimited membership) • Private channels (require invite links) • Public channels (searchable and open to anyone) • Chatbots • Viber Business Messages API • Stickers	Owned by Japanese tech conglomerate, Rakuten; sells in-app ads based on user metadata; some revenue from "Viber Out," a subscription service for calls.

Such a system prevents anyone aside from these parties—even the platform or service provider facilitating the delivery of the message—from deciphering the shared content.[13] For this reason, it is considered the gold standard in terms of privacy and security of communication.[14]

End-to-end encryption offers human rights defenders, political dissenters, and other activists the ability to communicate and organize without fear that their messages will be surveilled by repressive authorities.[15] But it also affords unscrupulous actors, including criminals, political extremists, and those behind disinformation campaigns, a useful means of concealing messages and remaining immune from the platform moderation that exists on more public platforms like Facebook and YouTube.[16]

Despite their vital role as channels for political communication, messaging apps are often overlooked in policy discussions concerning the impact of online platforms on information consumption. The major social media platforms—Facebook, X, YouTube and TikTok—still attract most of the attention as they are collectively seen to comprise the digital public square: the virtual sphere where members of the public exchange opinions and information.[17] Messaging apps, by contrast, are perceived as largely private spaces for communication among friends, family, and other known contacts. But their role as political instruments should not be overlooked.

Given that 2024 is the "biggest election year in history,"[18] with people in more than 50 countries going to the polls, this report unpacks how certain messaging platforms are facilitating manipulation of the information environment by politicians and their operatives. The report draws on interviews with political strategists in 17 countries; a survey of over 4,500 messaging app users in nine countries; and insights from cryptographers, cybersecurity professionals, and other experts to identify abuse strategies linked to specific platform features, designs, and policies. The report also offers actionable recommendations for messaging service providers, policymakers, and civil society to help counter propaganda efforts during upcoming elections, without undermining the privacy guarantees of end-to-end encryption or hampering pro-democratic activity.

Enormous reach

In many parts of the world encrypted messaging platforms have surpassed social media (i.e., Facebook, X, and TikTok) as the most widely used app category.[19] They continue to rise in popularity, especially in the Global South, in part because they enable free or very cheap and reliable communication across borders. In the U.S., usage of these apps is on the rise, particularly among diaspora communities.[20] WhatsApp's U.S. user base, for instance, grew by 9% in 2023 to roughly 100 million users.[21]

Among messaging platforms, WhatsApp is the largest, with more than two billion users worldwide. It is followed by Facebook Messenger (almost 1 billion), Telegram (over 900 million), and Viber (roughly 800 million). Another encrypted chat app, Signal, has a more modest user base of around 40 million but is becoming increasingly influential due to its robust encryption protocol and credible privacy guarantees.[22] These platforms differ in terms of their geographic reach, extent of encryption implementation, app features, and business models (see table on page 5).[23]

The messaging app market has exploded in recent years, enabled by the expansion of cellphone and Internet penetration in many parts of the world. Although several other messaging services exist,[24] this report will primarily cover four platforms—WhatsApp, Telegram, Signal and Viber. First, each of these apps applies end-to-end encryption to some extent—albeit minimally in the case of Telegram.

Second, there is some evidence of exploitation of each of these platforms by political propagandists during elections. We define political propagandists as individuals or groups working to leverage media and communication in purposeful efforts to manipulate public opinion, particularly during elections and other events of civic significance. Third, these platforms all have large global user bases. Although Facebook Messenger has significantly more users than Signal, Messenger started rolling out end-to-end encryption only in December 2023,[25] whereas Signal has always been—and continues to be—the purest example of an encrypted messaging service.

In fact, with the exception of Signal, it is imprecise to characterize these platforms as "encrypted messaging apps." The apps are encrypted to different degrees, and are much more than just messaging services. Telegram, in particular, can hardly be characterized as an encrypted chat app as large swaths of service are not end-to-end encrypted (only the "secret chats," which users have to manually enable and are only available for one-to-one conversations, are protected with end-to-end encryption).[26] Telegram users' sense of privacy[27] could be explained by the company's hands-off approach to moderation, coupled with its misleading use of the term "encryption" when promoting its service.[28] Recently, Telegram escalated its deceptive marketing, seizing on a politically-motivated and conspiracy-fueled crusade against Signal, to claim that Telegram is "the only popular method of communication that is verifiably private"—an assertion that contradicts the broad consensus in the cryptography community.[29] Telegram did not respond to requests for comment.

WhatsApp is mostly end-to-end encrypted,[30] but the app has grown from a simple messaging service to more of a one-stop-shop for digital communication.[31] The same applies to Viber,

whose features, ranging from "communities" of unlimited membership to "public channels," match the scale and openness of social media. WhatsApp and Viber, without employing a level of deception on par with Telegram's, also make blanket claims of security that are exaggerated and confusing to users.[32] While attracting consumers with promises of privacy, these apps subsist largely on revenue from subscription services, paid premium or business features, and advertisements—all of which involve the creation of tools that facilitate propagandists' exploitation of encrypted messaging.

What We Mean By 'Propagandist' and Other Key Terms

- We define **political propagandists** as individuals or groups working to leverage media and communication in purposeful efforts to manipulate public opinion, particularly during elections and other events of civic significance.[1]

- **Propaganda** in this report refers to attempts to influence a target audience through content that is **false or misleading**, and/or by employing tactics that are **manipulative or inauthentic**.[2] Some propaganda may be **disinformation**, which is false information disseminated with the intention to deceive, or **misinformation**, false information disseminated without an intent to deceive.[3]

- **False content** is content which is made up and refutable through fact-checking. For example, during the 2024 election cycle, the BJP in India falsely claimed that the main opposition party had pledged to take money from Indians and distribute it to Muslims, in an effort to rile up anti-Muslim sentiment in the country.[4]

- **Misleading content** is content that contains a grain of truth but is embellished with untrue or difficult to prove information. For example, Tunisian President Kais Saied claimed that sub-Saharan African migrants were to blame for the scarcity and unaffordability of bread. This narrative relied on the fact that sub-Saharan migrants were passing through Tunisia on their way to Europe (and the related likelihood that they were also buying bread) to suggest that the migrants were responsible for the "bread-crises," which were in fact due to economic mismanagement and other factors.[5]

- **Manipulative tactics** often involve appeals to identity, historical grievances, existing societal fissures, or other topics likely to elicit emotional responses. The BJP's strategy of stoking fear of Muslims among its Hindu base by promoting conspiracy theories about Muslims' alleged plans to take over India, for example, counts as a manipulative tactic. A manipulative tactic on messaging apps can also entail using networks of fake accounts to create the impression of popularity or public approval.

- **Inauthentic**, **or deceptive, tactics** on online platforms involve the creation of **fake accounts or groups** whose names conceal the true identity or purpose of the account holders. In Nigeria, propagandists have impersonated "Nollywood" actors to attract large followings for their fake accounts and channels.[6]

---

1 Woolley, S. (2023). *Manufacturing Consensus: Understanding Propaganda in the Era of Automation and Anonymity*, Yale University Press.

2 Academic literature sometimes refers to this as "negative propaganda," as differentiated from the neutral term, "propaganda," which is merely content created with the intention to change people's minds. See Lippmann W. (2018), *Die öffentliche Meinung: Wie sie entsteht und manipuliert wird*, Westend Verlag GmbH.

3 Information can also be true but placed out of context to create a false narrative, or "malinformation." See Wardle, C. & Derakhshan, H. (2017), Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making, Council of Europe. Given this report's spotlight on negative propaganda, our examples tend to involve disinformation, rather than misinformation. However, while this differentiation is helpful for analytic rigor, the intent cannot always be determined or the classification of content as misleading is often disputed. These limitations lie in the nature of any scholarly work in the social sciences which, while aiming to be objective, can never be separated by the individuals and their environments conducting the research.

4 https://www.thehindu.com/elections/lok-sabha/fact-checking-the-muslim-league-manifesto-claim-against-congress/article68110423.ece

5 https://timep.org/2024/07/11/rhetoric-and-repression-anti-migrant-discourse-as-a-political-weapon-in-tunisia/

5 Interviews in Nigeria in January and February 2024 (online).

2. How Messaging Platforms are Being Exploited in Elections

In-depth interviews with political operatives in 17 countries reveal that propagandists regularly exploit the features and offerings of messaging platforms before and during elections in attempts to shape public opinion.

> **Many messaging platforms allow for the creation of large groups, which plays directly into political propagandists' hands.**

Their main tactics involve using in-app broadcasting features to promote content virality; exploiting paid business accounts and premium subscriptions to reach even larger audiences; and scaling outreach and engagement by means of fake or "sock puppet" accounts. This section will explain some of the mechanics behind these strategies.

Broadcasting toolkit

Building on years of experience using encrypted chat apps during elections, propagandists have developed a "broadcasting toolkit."[33] This toolkit transcends the in-app "broadcast" feature, which allows for simultaneous messaging to several contacts.[34] It involves combining the platforms' various communication tools in sophisticated ways to achieve the viral dissemination of content while capitalizing on the intimacy and directness intrinsic to private messaging. As a political consultant in India put it, "Perfect[ing] the broadcasting toolkit (…) is the backbone of any campaign."[35]

The first step in propagandists' broadcasting strategy involves curating systems of distribution. These are networks of groups, constructed over the course of months or years leading up to elections, which are carefully calculated to maximize engagement and speed of dissemination. In practice, propagandists build these networks by infiltrating preexisting groups as well as creating new ones.[36] Preexisting groups offer the advantage of being organically—or spontaneously—formed. Even when the groups are originally apolitical, propagandists can exploit members' professed interests to craft political messages that are likely to resonate. Constructing new groups is more laborious—and often involves collecting individual phone numbers by going door to door, or working with data brokers to curate lists—but gives propagandists greater control over the narrative and packaging of information, for example by claiming to represent "official" party groups.[37]

Many messaging platforms allow for the creation of large groups, which plays directly into political propagandists' hands. For example, Viber's lack of member limits in its "communities" and "channels" has enabled the Filipino government to distribute aggressively pro-government messages directly to millions of users on an app that Filipino residents use

on an hourly or daily basis.[38] These messages have ranged from HBD (Happy Birthday) messages and stickers celebrating certain political figures, to hawkish messages praising Duterte's draconian "war on drugs" and smear campaigns against people criticizing those harsh measures, declaring them "enemies of the state."[39] According to a local opposition politician, the Duterte government "invaded Viber" during its administration. And this was a strategic decision—according to our survey of messaging app users, 26% of Filipinos use messaging apps (of which Viber is among the most popular) several times per hour and 43% use them several times per day.

WhatsApp's decision in 2022 to increase the maximum group chat size to 1,024 members as well as to roll out supergroups, called "communities," moved it closer to Viber in terms of the app's utility for mass messaging.[40] In the words of one political consultant who works primarily with the BJP in Uttar Pradesh and Himachal Pradesh, "the more people we can reach with one message, the better."[41] WhatsApp, cognizant of its responsibility to counter exploitation, recently ramped up its detection and removal of inorganic group creation—and these efforts were reflected in propagandists' experiences in 2024. The same political consultant who had benefited from large group creation in previous elections reflected: "We realized WhatsApp did not like our automated attempts. It was convenient for us, but I guess they found it suspicious."[42] So, this election season, propagandists in India changed their tactic: Instead of creating large groups from scratch in a short period of time, they gradually and methodically added new phone numbers into existing groups.[43] The new members were a mix of regular voters, who were more likely to remain engaged in these pre-existing groups, and other propagandists, who could simulate organic engagement by taking turns circulating and supporting political content.[44]

In Maharasthra, the second largest state in terms of representation in the lower house (Lok Sabha) of India's parliament and home to the commercial hubs of Mumbai and Pune, smaller parties like the Nationalist Congress Party (NCP) have been building WhatsApp communication and distribution networks for years, rivaling those of the BJP. These efforts potentially helped the surprise downfall of the BJP and respective gains of the INDIA alliance (which the NCP had joined and partially stayed with) in Maharashtra. However, the BJP also benefitted from smaller local outfits helping their WhatsApp campaigning. For example, the Maharashtra Navnirman Sena (MNS) Party was prolific in its outreach as the below poster shows. The data the MNS collected was used for campaigning by Modi's coalition.[45]

The ability to form numerous interconnected groups on messaging apps facilitates coordination among political operatives and their targeted messaging. This cross-platform strategy was described by an IT specialist for the BJP in India as enabling a highly organized inner-party system where "there's various levels of these groups. There's headquarters..there's state level groups, there's district level groups, and finally there's booth level groups, as in booth-poll-voting level groups, and these are the [groups] that are usually public. The rest [of the groups] are usually private, [for] party communication."[46]

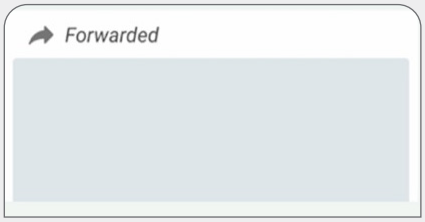
After setting up networks of distribution, propagandists are ready to deploy their broadcasting toolkit. In 2024, the main app features comprising this toolkit have been: intra-app forwarding capabilities, cross-posting, status updates or "stories," channel feedback loops, and bots.

Deceptive recruitment

The text in large font says, "I'm a protector of Hinduism and a worker for Maharashtra." A QR code on the right side is accompanied by the text, "this is the place to get yourself voter registration." This part is deceptive. The QR code links to the website of a politician for the MNS Party, a far-right Hindu Nationalist Party. The site asks visitors to submit their phone number. This is a common way for political organizers to recruit people into WhatsApp groups.

Photo by Inga K. Trauthig.

In 2018, WhatsApp implemented forwarding labels and limits in an attempt to "slow down the spread of rumors, viral messages, and fake news."[1]



These messages can be forwarded to up to 5 chats at a time

To help keep conversations personal, these messages can be forwarded to 1 chat at a time

[1] https://faq.whatsapp.com/1053543185312573

Intra-app forwarding is the oldest and most widespread method of content dissemination on messaging platforms. Rather than having to copy and paste messages from one chat to another, the forwarding feature enables the instantaneous sharing of content with many contacts at once. It is a powerful, if rudimentary, tool. In recognition of its role in the rapid spread of false information during the 2018 Indian elections, WhatsApp implemented forwarding labels and limits.[47] Since 2018, any message that has been forwarded at least once carries a tag at the top with a single arrow and the term "forwarded," and any message that has been sequentially forwarded five or more times carries a double-arrow icon and the phrase "forwarded many times." In addition, messages labeled as "forwarded" can only be sent to up to five additional chats at a time, whereas "frequently forwarded" messages can be sent to only one chat at a time.[48]

There is some evidence that these measures by WhatsApp, while laudable, have had a limited impact in India and elsewhere. Our survey of messaging app users in 2024 indicates that users typically do not notice which messages have been forwarded.[49] More concerning, a study from 2023 found that forwarding tags were ineffective at stopping the circulation of misinformation because some WhatsApp users misinterpreted the tags as signaling important content, and others failed to draw the connection between forwards and potential disinformation.[50] Another study, from 2019, found that early rollouts of WhatsApp's forwarding limits in Brazil and India merely slowed the spread of information; the limits did not effectively block disinformation campaigns in public groups when content was highly viral.[51] It is likely that WhatsApp's decision to increase maximum group sizes in 2022 further eroded the impact of forwarding limits on content virality. Nevertheless, WhatsApp deserves some credit for attempting to curb abuse of one of its virality-promoting features.

Other messaging platforms have not been as conscientious. Viber does not have any limitations on message forwarding.[52] Telegram also does not limit forwarding in any way and even facilitates its automation with a dedicated "forwarding bot" feature.[53] This type of automation can amplify propaganda while simultaneously suppressing other types of content.[54] In the leadup to the U.S. 2024 presidential elections, Telegram's forwarding tool played a role in the spread of disinformation about President Biden among Latino communities in South Florida.[55] In Hungary, before the 2024 EU elections, Telegram's forwarding bot was deployed to carry out anti-LGBTQ smear campaigns and attacks against pro-democracy civil society organizations portrayed as Western-controlled.[56]

Another form of forwarding, called "cross-posting," helps propagandists disseminate content across platforms, thereby transcending the walls of each app ecosystem and reaching even larger audiences.[57] Some apps make such inter-platform forwarding seamless through dedicated tools. On Telegram, users can develop cross-posting bots that automate the sharing of content from Telegram to X (formerly Twitter), creating a disinformation feedback loop among the two platforms. In India, another social media platform, ShareChat, often serves as the hotbed of manipulated content that ends up spreading virally on WhatsApp, thanks in part to an in-app feature that allows for cross-posting to WhatsApp with one click. Material that reaches WhatsApp can then be shared easily —and, as of April 2024, automatically —with audiences on other platforms in the Meta family of apps, including Instagram and Facebook.[58]

Due to many messaging apps' tendency toward feature bloat,[59] propagandists now have several other tools in their broadcasting toolkit beside forwarding and cross-posting. Stories and channels are two app features that provide propagandists with easy and instantaneous access to potentially large audiences. Stories is a feature native to traditional social media platforms like Instagram and Snapchat, but one that has now been integrated into many

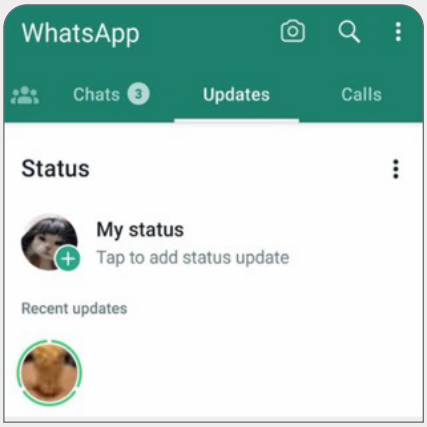
messaging platforms (on WhatsApp they are called “status updates”). On messaging apps, stories work in the same way as on social media—by allowing users to passively broadcast content, often photos and memes, to all their contacts for a certain period of time. Political propagandists capitalize on this feature not only by posting stories themselves but also by encouraging supporters to repost a piece of content to their stories (in Telegram’s jargon, “repost to your story”), thereby multiplying the broadcasting effect.[60]

Channels, which are one-way mass broadcasting media available on Telegram, Viber and (since November 2023) WhatsApp, can serve as the origin of false, misleading or manipulative information which then circulates on groups and private chats. In the weeks leading up to the 2024 polls in India, “official” BJP WhatsApp channels shared videos with heavy anti-Muslim rhetoric, calling on those who want to “save India from the Muslims” to vote for Modi. As the content on channels is not encrypted, WhatsApp was able to take down the posts a few hours after they were shared—but by then the content had spread to the WhatsApp ecosystem of encrypted groups and private chats.[61]

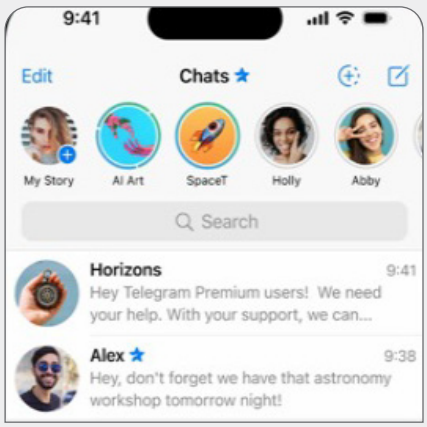
In addition to capitalizing on these virality-promoting features individually, propagandists combine channels, stories, and forwarding tools to create feedback loops that extend the content’s longevity. Essentially, to increase the chances that a particular disinformation narrative will gain traction, or “trend,” propagandists work to ensure that the same content continues popping up in different parts of the platform ecosystem. A piece of misleading content emerging from a channel may be forwarded to a group chat, where some members may then be encouraged to repost it to their story.[62] Any message forwarded from a channel displays a note below that says “view channel,” which is an effective recruitment tool to bring additional users into the channels, completing the feedback loop.[63]

Mexico provides a case in point. In the last election, propagandists created a copycat version of Animal Politico, a reputable Mexican political news outlet, which they managed to get verified by WhatsApp. Upon setting up the fake channel, they routinely posted misleading material and coordinated engagement with that material on the channel. Finally, they forwarded the content from the channel to other group chats and channels, where the “view channel” tag attracted attention to the content and drove additional people into the fake Animal Politico channel, thereby reinforcing the loop.[64] In this instance, the channel was a strategic coup for propagandists who then used the channel’s success as a selling point when promoting their services to local politicians. Unfortunately for them, the channel was disbanded, likely after having been reported to WhatsApp.[65]

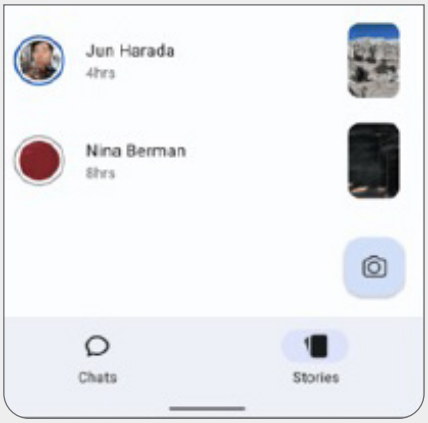
Comparing the “stories” feature on WhatsApp (left), Telegram (center), and Signal (right). Viber does not currently have a stories feature



Stories, called “status updates” on Whatsapp, are featured in the “Updates” tab under “Recent Updates.”



Stories on Telegram are featured at the top of the “Chats” tab with profile rings that users can turn on.



Stories on Signal are featured in the “Stories” tab.

Faux businesses

With the broadcasting toolkit, propagandists can promote content virality across their curated distribution lists. But political operatives can reach even larger—and potentially unlimited—audiences within a faster timeline by paying for premium features. In the case of WhatsApp and Viber, this enhanced access is (nominally) available only to commercial entities. In the case of Telegram, anyone who pays for a premium subscription can reap similar benefits.

In 2018, WhatsApp rolled out the WhatsApp Business Platform, also known as the WhatsApp Business API, enabling medium and large enterprises to reach potential customers on a global scale.[66] In addition to mass messaging features, the Business Platform confers a "green tick" verification to the account and provides tools for automated messaging. WhatsApp notes that it "prohibits use of the WhatsApp Business Platform (API) by political parties, political candidates, politicians, and political campaigns." Yet company representatives acknowledged in an interview that "the Business Platform is available for government use," which complicates the distinction between government communication and campaign messages sent by an incumbent. Perhaps this is how, in March 2024, BJP operatives were able to broadcast political propaganda to non-consenting users under the guise of a verified business account called "Viksit Bharat Sampark."[67]

According to WhatsApp representatives, the platform "prevent[s] political abuse of our paid messaging service through automated detection, integrity gating during onboarding, user reports, and manual human review that assess business account information to ensure compliance with our government and political use policy restrictions." They added: "If a political actor tries to circumvent our systems to access the Business Platform, we take appropriate action to enforce policy violations," and "where we have evidence of direct campaigning, advocacy, or promotion related to an election—these will be actioned in accordance with our policies." Furthermore, the company noted, "[d]uring electoral periods, we devote additional resources to these compliance efforts to ensure adherence to our policies."

Beyond policies and their enforcement, WhatsApp representatives highlighted the Business Platform's anti-abuse measures baked into its design. For example, they noted, "businesses must obtain affirmative opt-in and permission from users before sending messages on the Business Platform, which helps prevent unwanted outreach." Furthermore, the ability to send messages to an unlimited number of users is contingent on a business getting a high messaging quality rating. In addition, WhatsApp requires that the first message sent by a business to a new user through the API be pre-approved and conform with a template for business-initiated conversations.[68]

On paper, these requirements serve as strong safeguards against abuse of the Business Platform. In practice, however, political propagandists report being able to game WhatsApp's verification system by submitting account requests using fake business websites or credentials.[69] In Nigeria, for example, propagandists have used several tactics to get away with political messaging via business accounts. One tactic is to get verified on X (formerly Twitter) with a fake business name and use that as proof of existence when registering for WhatsApp's API. A second is to impersonate "Nollywood" actors and then obtain verification as a business account. A third is to register their number as a business account hoping that WhatsApp accepts it.[70]

Viber's Business Messages API, unlike WhatsApp's, is set up such that business entities must obtain customer consent to receive messages. This

> **"Our survey of messaging app users confirms that propagandists are able to reach non-consenting recipients despite existing safeguards. Among users who reported receiving political content via messaging apps in the last year, 55% said some of that content came from people or accounts they did not recognize."**

consent can be acquired through links shared via email or text messages.[71] This opt-in system, while adding some friction to political propagandists' outreach, has loopholes as well. In Ukraine, where Viber has near-100% penetration, political consultants have managed to obtain verified Viber business accounts under false pretenses via one of Viber's many official messaging "partners" or vendors. Then, to secure consent from unwitting constituents, operatives launched campaigns on social media and physical billboards with QR codes coaxing users into subscribing to their mailing lists.[72]

Viber representatives note that "[p]er the Viber Advertising Policy, Political ads are Restricted Content. All advertisers and users of our API are expected to adhere to the Viber Advertising Policy." Yet, instead of monitoring business accounts after having approved them to ensure compliance with policies, Viber representatives said they "do not monitor content," leaving the door open for indefinite policy infringement. In the Philippines, ads publicizing the daughter of former Philippines president Rodrigo

Duterte, Sara Duterte-Carpio, were widespread on Viber in the 2022 presidential election.[73]

Our survey of messaging app users confirms that propagandists are able to reach non-consenting recipients despite existing safeguards. Among users who reported receiving political content via messaging apps in the last year, 55% said some of that content came from people or accounts they did not recognize.[74] The fact that more than half of messaging app users receive political content from strangers belies the purported role of messaging platforms as spaces for private communication among family and friends.

The shifting character of these apps can be explained by the platforms' thirst for revenue. According to unofficial industry analysts, WhatsApp derives more than one-third of its revenue from business features.[75] For Viber, the share of revenue from business accounts is much smaller, in part due to its monetization of in-app ads.[76]

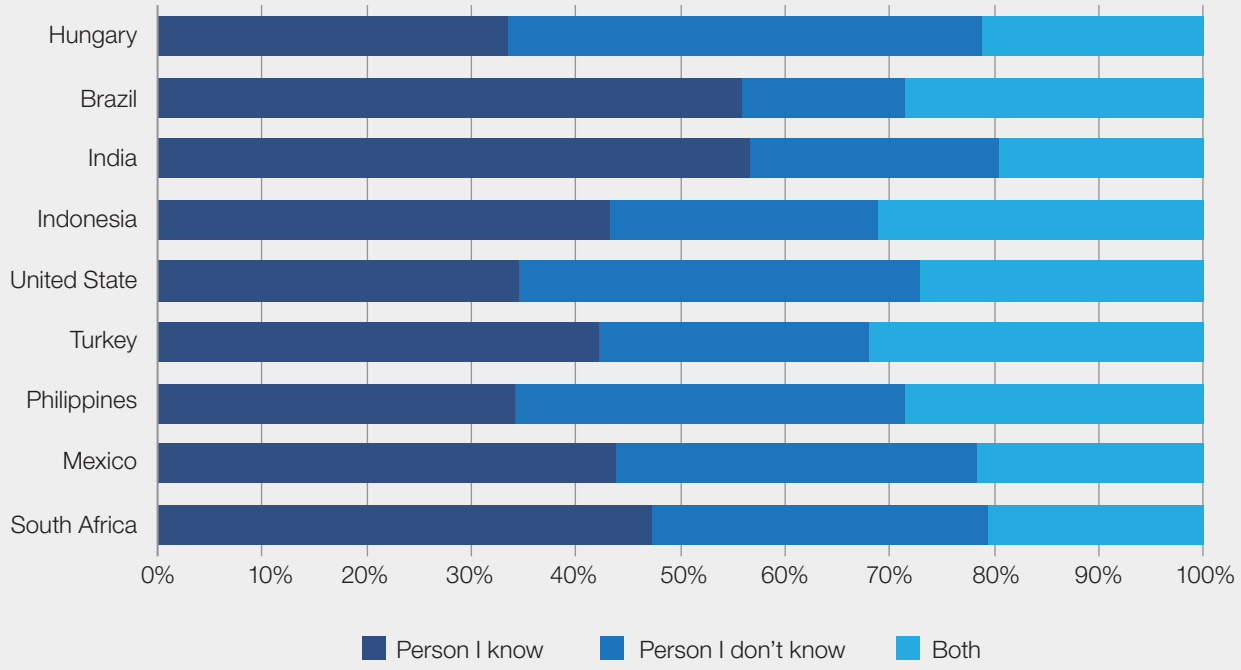
Telegram, a private company reportedly considering an initial public offering, has faced pressure to prove the viability of its business model.[77] Its rollout of premium subscriptions and advertisements is part of an effort to make a business case to investors. Premium subscribers, who pay $4.99 per month, have access to a series of perks including additional accounts, premium stickers, advanced chat management options, profile badges and, as of March 2024, a suite of business features—regardless of whether or not they are actually businesses.[78] With access to premium badges, political propagandists are able to masquerade as "official" accounts without having to attain verification badges,[79] which are nevertheless possible to obtain through Telegram's relatively relaxed verification process.[80] Furthermore, with access to business features, they can scale their messaging operations using a collection of automation and customization tools.[81]

Telegram's Ad Platform, which generates an increasingly important portion of the company's revenue stream, allows anyone to buy short-form ad placements in channels with high subscriber numbers.[82] Although according to Telegram's policies, ads "must not promote political campaigns, elections, political parties, candidates, political or religious movements,"[83] political propagandists report being able to place ads on highly frequented news channels for their own election-related channels and groups.[84]

Sock puppets

For years, political propagandists have operated fake, or "sock puppet," accounts on messaging apps to scale their operations and intensify their deception.[85] Such accounts purport to represent a person or entity with a particular viewpoint, but their personas are run by entities looking to "launder" information to make it seem more legitimate or trustworthy.[86]

Origin of political content on messaging apps

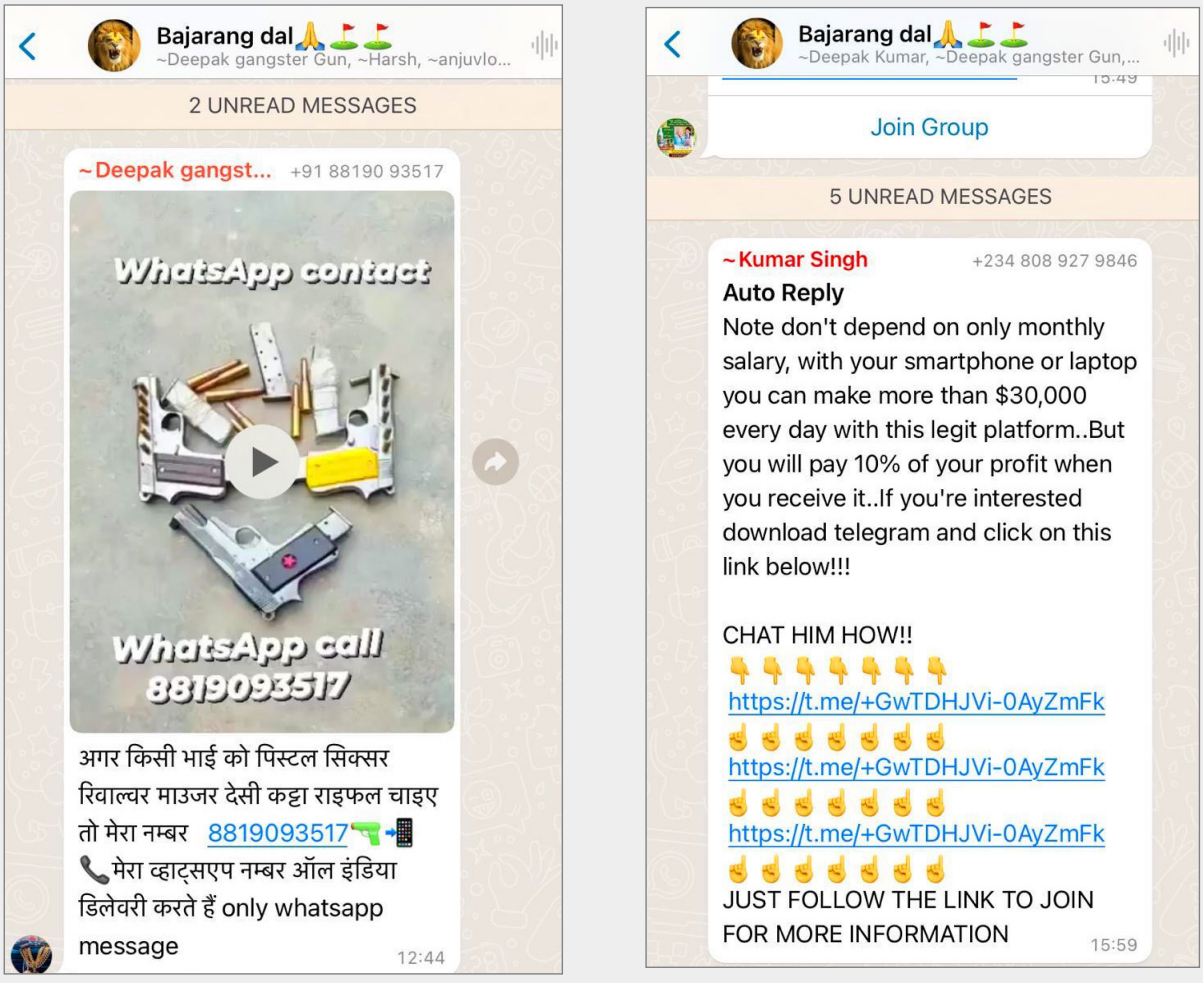


In Bolivia, a legal advisor to the legislature is anxious about the anticipated national elections next year because "WhatsApp is uncontrollable." She elaborated: "Every bureau of the government has a WhatsApp group or channel for spreading information (…) [but] bad people pretend to be the government and create hateful rumors (…)."[87] One group that purported to be from the Ministry of Labor, Employment, and Social Security created polarizing rumors based on false claims that the government unfairly favored indigenous populations of Bolivia by giving them higher social benefits than was the case.

The use of sock puppet accounts is not unique to messaging platforms and characterizes covert influence campaigns across social media.[88] On messaging platforms, sock puppets are arguably more problematic because they can operate with greater obscurity while benefiting from the directness of communication that the platforms provide. With the exception of Signal, messaging platforms have failed to enact sufficiently robust account creation restrictions, perhaps because placing such limits would diminish user numbers and, by extension, their appeal among investors and advertisers.[89]

The most sophisticated and well-resourced political propaganda campaigns arm their operatives with multiple phones and computers from which to operate numerous online accounts.[90] Accounts are often developed over time and deployed for the purposes of manipulation during particular events or circumstances, in order to grant the illusion of legitimacy. But messaging platforms may have made investments in phone farms less necessary by allowing users to manage multiple accounts from a single device.

Screenshots from the extremist Bajarang Dal WhatsApp group in India, advertising weapons for sale and linking to Telegram for business opportunities.



Screenshots by Inga K. Trauthig.

Telegram is, again, the most permissive. Through the Fragment blockchain, managed by Telegram itself, users can actually buy phone numbers and Telegram usernames.[91] Moreover, Telegram is piloting the ability of users to manage unlimited accounts—a feature on the exploratory Telegram X platform, available in some countries.[92] These affordances will likely make the operation of fake accounts and phone farms significantly more widespread.

WhatsApp and Viber, by contrast, allow the creation of up to two accounts per phone number. But propagandists said they find ways to circumvent these limitations. One tactic involves registering different accounts on the desktop and phone versions of the app.[93] Another consists of taking over phone numbers that are no longer in use and broadcasting propaganda via status updates.[94] In some cases, political parties in India have bought phone numbers from Telecom companies.[95]

The ability to create multiple accounts fuels the market for professional political trolls—content producers who specialize in creating partisan material that is highly manipulative and incendiary in nature.[96] Such trolls exploit the relative obscurity of encrypted messaging to engage in illegal, or borderline illegal, behavior. For example, they might violate election laws, such as requirements to label political advertisements and observe pre-election silence, while flying under the radar of electoral authorities. In India, according to an investigation by *Rest of World*, the BJP has leveraged an army of volunteers at the national, regional, and district levels to manage WhatsApp groups and disseminate content via third-party accounts. Because the accounts, and the WhatsApp groups they manage, are not part of the "official" BJP communications apparatus, the party has given them license to spread inflammatory messages calculated to exploit specific grievances in the population.[97] Some third-party accounts have then used the apps' "direct messaging" functions to recruit the most active and loyal supporters to militant organizations.[98]

Such practices are an outgrowth of similar efforts made by governments and other powerful political groups to exert influence via trolling on platforms like Facebook, X, and YouTube. These state-sponsored trolling campaigns leverage inter-platform communication across a variety of digital channels, with varying levels of government or party involvement. Messaging apps have been integrated into the political trolling toolkit.[99]

Another propaganda tactic involves the use of bots. The growing presence of chatbots, including those powered by generative AI, on messaging apps further erodes the assumption that these apps are designed for communication among known contacts. On Telegram, bots can be programmed to carry out a plethora of tasks, from managing a channel sign-up and consent process to receiving donations and sending out automatic updates to channel subscribers based on news developments.[100] Telegram hosts over 10 million free bots created by third party developers, and these bots can be integrated into other chat apps including WhatsApp.[101]

On Viber, until May 2024, users were able to create and publish their own chatbots for free, and could even send up to 10,000 chatbot-initiated messages per month, with a cost-per-message charged beyond this limit.[102] In the Philippines, during the last round of presidential elections, propagandists created Viber bots that could proliferate text and video messages quickly and widely. Viber points out that chatbots created after May 2, 2024, are covered under the company's new chatbot commercial model, whereby chatbot owners must pay a 100-euro monthly maintenance fee and are charged a one cent cost-per-message for each chatbot-initiated message. Once a user replies to this chatbot message, they have a 24-hour free messaging session.

> **"On messaging platforms, sock puppets are arguably more problematic because they can operate with greater obscurity while benefiting from the directness of communication that the platforms provide."**

Chatbot owners are also given a free welcome message with which they can begin a conversation.[103] Like the old model, this chatbot commercial system is susceptible to misuse by political actors. Chatbot owners can use the free welcome message or the one-cent chatbot initiated message to spread political propaganda to users, and once a user engages with this message, all messages are free for the next 24 hours. Moreover, it is not always clear to users whether they are interacting with bot or human users as not all bots are labeled as such.

Beyond aiding propagandists in scaling information distribution, in a way similar to manually-run sock puppet accounts, bots are increasingly taking over content production. WhatsApp's recently launched AI-powered chatbot, which is featured prominently at the top of the app, can be prompted to answer questions about real-world events, compose "news" stories, draft posts simulating communications from specific political candidates and, in some countries, insert AI-generated images directly into chats.[104] Political strategists are using these tools to produce synthetic content targeted at different audiences.[105] Generally, these tools act as accelerators of human-driven strategies, and their full impact on elections is yet to be seen.

Extremists' Use of Messaging Apps to Mobilize for Violence

Extremist actors, including terrorist groups, are known to exploit online spaces that provide refuge from platform moderation. These relative safe havens often act as venues where the most dangerous and radical activity, such as mobilization for violent action, occurs.

Research into extremist activity on encrypted messaging platforms is scant because of the challenge of conducting ethical research on platforms designed for privacy. But open-source investigations by Tech Against Terrorism, a United Nations-backed initiative, reveal that the average removal rate of terrorist content is lower on messaging apps than on almost all other platform types, including social media, video-sharing sites, and file-sharing platforms.[1]

Other research suggests that extremist groups do not actually need the protection of end-to-end encryption to carry out illegal activity because they already find safe harbor on (mostly) non-encrypted platforms like Telegram and Discord.[2] According to the executive director of Tech Against Terrorist, "Telegram is fundamental to the terrorist content ecosystem,"[3] even though most of the app is not end-to-end encrypted and the platform could moderate content if it wanted to.[4] The platform's CEO and founder, Pavel Durov, a Russian exile who describes himself as a die-hard free speech libertarian,[5] has in most cases refused to block channels disseminating extremist content.[6]

Evidence of illegal activity on Telegram is plentiful[7]—but it is also relatively easy to find because a large part of the platform is public. By contrast, there is little publicly available evidence of extremist activity on truly private platforms like Signal. Aside from court documents pointing to the use of Signal by organizers of the January 6th insurrection at the U.S. Capitol,[8] there is little open-source evidence that Signal is widely used by terrorists and other criminal sectors.[9] But, even if there were such activity, Signal's position is that it would not be able to do anything because "we would not know."[10]

---

1 Hadley, A. & Maguire, E. (2024, April 30). *Combatting Terrorist Misuse of Messaging Apps* [Webinar]. Tech Against Terrorism. https://techagainst-terrorism.org/events/webinar-series-combatting-terrorist-misuse-of-messaging-apps

2 Interview with Hannah Rose on March 15, 2024. For evidence of extremist activity on Discord, see Olaizola Rosenblat, M. (2023, May 16). Gaming the system: How extremists exploit gaming sites and what can be done to counter them. *NYU Stern Center for Business & Human Rights*. https://bhr.stern.nyu.edu/publication/gaming-the-system-how-extremists-exploit-gaming-sites-and-what-can-be-done-to-counter-them/

3 https://techagainstterrorism.org/in-the-news/graphic-videos-of-hamas-attacks-spread-on-x-0-0-0. See also, https://www.ft.com/content/c70ef7d6-230a-4404-b854-2e75fe0f2e0a ("Telegram is social media for organised criminals").

4 Hadley, A. & Maguire, E. (2024, April 30). *Combatting Terrorist Misuse of Messaging Apps* [Webinar]. Tech Against Terrorism. https://techagainst-terrorism.org/events/webinar-series-combatting-terrorist-misuse-of-messaging-apps

5 https://www.ft.com/content/c70ef7d6-230a-4404-b854-2e75fe0f2e0a

6 https://www.wired.com/story/telegram-hamas-israel-conflict/

7 See, e.g., https://www.wired.com/story/telegram-hamas-channels-deplatform/#:~:text=A%20WIRED%20investigation%20reveals%20that,but%20they%20are%20still%20there.; https://www.ft.com/content/c70ef7d6-230a-4404-b854-2e75fe0f2e0a; https://www.washingtonpost.com/politics/2022/04/23/telegram-platform-right-wing/

8 https://iapp.org/news/a/investigators-used-encrypted-signal-messages-to-charge-capitol-riot-defendants

9 Interview with open-source intelligence analyst at Tech Against Terrorism.

10 https://www.theverge.com/23409716/signal-encryption-messaging-sms-meredith-whittaker-imessage-whatsapp-china

3. Conclusion and Recommendations

Messaging apps provide the infrastructure for political manipulation campaigns that promote deliberately false or misleading information. While presenting themselves as platforms designed for secure and private communications among loved ones, some messaging apps monetize their products through features that enable large-scale reach and message virality.

> “Among messaging app users who received political content from strangers, 52% said the content had significantly or somewhat influenced their opinions.”

Relying on these features, propagandists deploy tactics that sway a significant portion of constituents. According to our survey, 62% of messaging app users across nine countries received political content on those apps, and over half (55%) of that political content came from people or accounts which users did not know and did not choose to follow. Further, of those who received political content from strangers, 52% said the content had significantly or somewhat influenced their opinions.

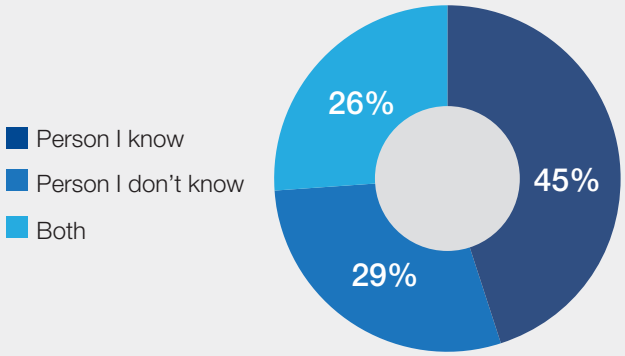
Moreover, platforms like WhatsApp, which apply end-to-end encryption by default to most features, lack many of the traditional mechanisms associated with content moderation aimed at tracking and controlling abuse. Others, like Telegram, use the term “encryption” in a misleading way, perhaps to justify their lack of robust moderation of public information in the face of harmful, and sometimes illegal, activity on the platform.

Simultaneously, encrypted communication is a tremendous asset for pro-democracy activists working in limited media systems and societies increasingly facing the pressure of surveillance capitalism and other attacks on privacy. These considerations make it clear that platforms must not “break,” or in any way undermine, encryption in an effort to mitigate political propaganda and other illicit behaviors on their apps.

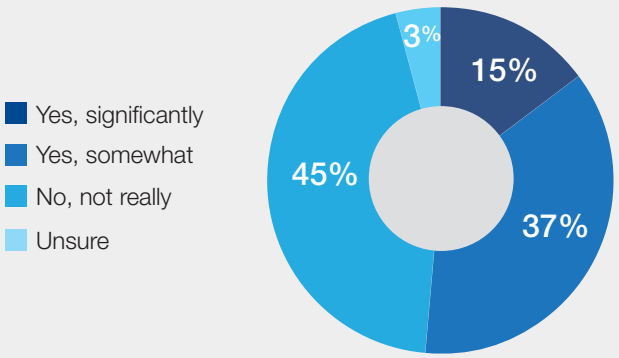
Fortunately, there are ways to curb misuses of messaging platforms while preserving the crucial democratic benefits of end-to-end encrypted information sharing. The recommendations below provide concrete steps that messaging services can take to safeguard their products from electoral manipulation. Policymakers, researchers, and other groups have important roles to play in supporting these efforts, and we make pragmatic suggestions for them too.

Provenance and influence of political content on messaging apps

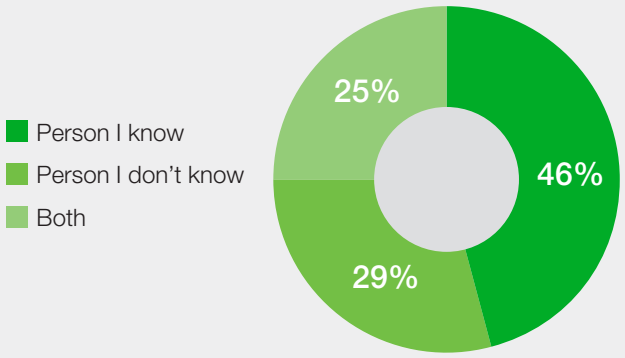
Origin of political content on messaging apps - total



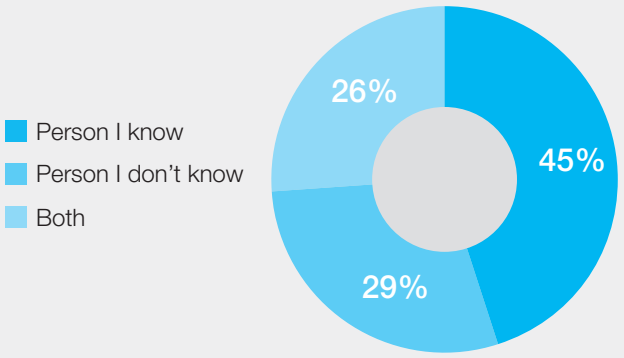
Do you believe political content shared on messaging apps has influenced your political opinions or beliefs?



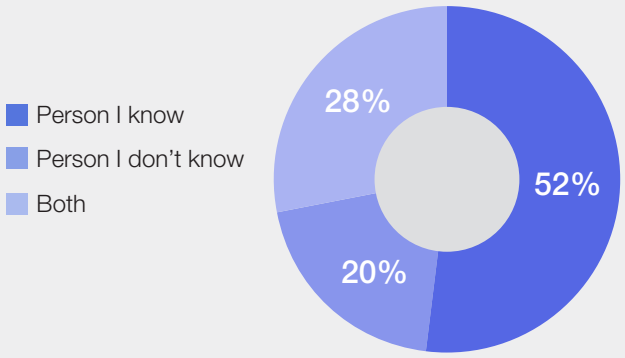
Origin of political content on Whatsapp



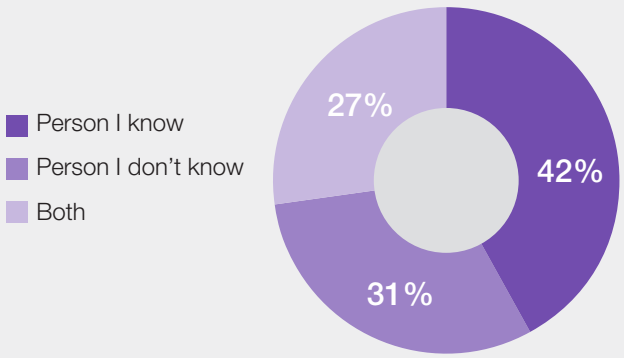
Origin of political content on Telegram



Origin of political content on Signal



Origin of political content on Viber



Recommendations

For messaging services

1 Establish strict account-creation limits and close technical loopholes to counter phone farms and covert influence networks.

To remain trustworthy channels for communication, messaging services need to crack down on coordinated, inauthentic activity. Phone farms and trolls dedicated to spreading manipulative content on messaging platforms rely on the ability to create and manage multiple accounts from a single device. One of the most direct ways for messaging services to curb political manipulation campaigns is by limiting the number of accounts someone can create and manage with a single device. Therefore, from the point of view of abuse mitigation, platforms should place limits of one account per device, as Signal does,[106] or up to two accounts if the platform offers business messaging services, as do WhatsApp and Viber. Telegram's policy of allowing the operation of multiple accounts from a single device seriously undermines the authenticity of communications on its service and facilitates its exploitation.

However, platforms should consult researchers who have extensively studied benevolent uses of the apps, and undertake empirical studies to determine an account limit that adequately balances the legitimate use of multiple accounts with the need to curb phone farms. Short of establishing absolute limits on the number of accounts per device, platforms should place limits on the pace at which new accounts can be created.[107]

Messaging services should also invest resources in closing technical loopholes that allow propagandists to bypass account limitations—for example, by ensuring that desktop and phone versions of the app are correctly synced, and combatting the malicious takeover of old phone numbers by prompting users to update their contacts before switching to a different number. WhatsApp helpfully flags when users have changed their phone number; other messaging apps should do the same.

2 Restrict large-scale broadcasting to verified channels and vetted business accounts.

Messaging services need to strengthen their vetting processes for channels and verified accounts, including business accounts, to ensure that inauthentic accounts do not benefit from the legitimacy and large-scale broadcasting opportunities conferred by such status. Field research indicates that propagandists are responsive to rigorous vetting processes. In India, several BJP and Congress officials said they refrained from using the WhatsApp Business Platform for fear that WhatsApp would detect misuse of the platform and close their other WhatsApp accounts.[108] The same deterrent has worked on Viber in some contexts, where some propagandists in Ukraine reported seeing their business accounts disabled after posting political content under false pretenses.[109] Nevertheless, some political campaigns in India and elsewhere continue to find loopholes in the business verification processes, which points to the need to further strengthen those systems.

3 Strengthen cross-industry and multi-stakeholder cooperation to identify inauthentic activity.

Messaging services should engage with peer platforms to identify cross-platform inauthentic activity. Political propagandists do not typically use just one platform when engaging in manipulation campaigns. They diversify their efforts across platforms, not only to build up insurance against potential moderation but also to play platforms' strengths (or weaknesses) against each other. Political actors in India have leveraged the platform ShareChat to stir up publicity around political stories before importing those narratives into WhatsApp.[110] If ShareChat and WhatsApp were to communicate findings of inter-platform inauthentic activity, they might have a higher chance of diminishing harmful disinformation before it goes viral.

Messaging apps are embedded in a propaganda ecosystem, so those companies should join existing efforts of cross-industry collaborations such as the Global Internet Forum to Counter Terrorism (GIFCT),[111] which currently only WhatsApp is part of. Further, institutionalized mechanisms for civil society actors to reach relevant representatives of the companies should be strengthened, especially in the Global South where the effects of false, misleading, and hateful content shared on messaging apps are most pronounced.

4 Support and improve access to accredited tiplines.

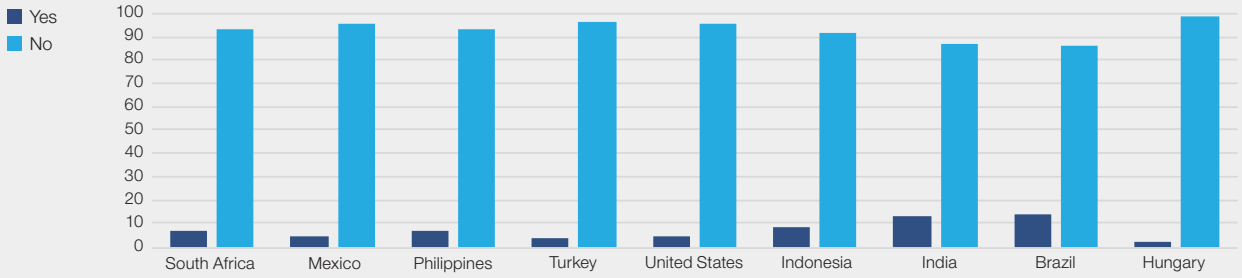
Platforms play a critical role in safeguarding their products from manipulation, but users are not powerless either. In fact, on end-to-end encrypted platforms, users are the parties best positioned to control their information diets. Researchers have proposed a number of tools for facilitating user-driven fact-checking on messaging apps, which platforms should consider implementing or improving. A large majority of messaging app users surveyed expressed a demand for such tools, and none of these affordances would violate the privacy and security guarantees of end-to-end encryption.

One approach involves supporting the operation of verified "tiplines." These are dedicated messaging app accounts managed by independent media or civil society organizations to which users can submit "tips" for fact-checking.[112] Platforms do not control tiplines, nor do they have access to the information exchanged. The decision of what to do with the results of the fact-checking process is entirely that of the user who submitted the evidence and of the organization that provides the fact-checking. But platforms can play a constructive role by authenticating tiplines, as they do with businesses and other verified accounts.

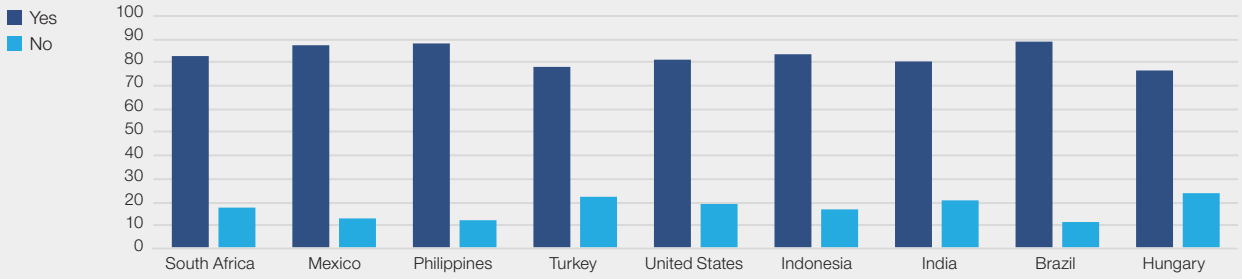
Among messaging apps, WhatsApp leads the way on tiplines, although there are troubling reports, which the company disputes, that Meta has cut funding for fact-checking on WhatsApp in the lead-up to the 2024 elections.[113] Since 2018, WhatsApp has partnered with a number of media and civil society organizations around the world to provide fact-checking services to its users through authenticated accounts. WhatsApp is transparent about its partnership with members of the International Fact-Checking Network (IFCN) which run those tiplines, and lists the organizations in a directory on its website.[114] Other messaging apps host tiplines but do not provide the same kind of authentication and support.

Even WhatsApp, however, can improve the accessibility of its tiplines through user interface improvements.[115] Currently, few users know about or use tiplines, and the cumbersome way in which tiplines are accessed is part of the problem.[116] According to our survey, only 7% of app users across the nine countries said they had ever contacted a tipline, although 83.4% said they would find such an option useful.

Have you ever contacted a tipline or fact-checking organization to check the truthfulness of a piece of political content received on a messaging app?



Do you think it would be useful to be able to contact a fact-checking organization to check the truthfulness of a piece of political content received on a messaging app?



5 Empower users to limit their exposure to manipulated content and disinformation through dedicated in-app affordances.

There are other forms of user-driven fact-checking that platforms can directly support through dedicated in-app affordances. One idea, suggested by Kiran Garimella, a professor of information science at Rutgers University, is for messaging apps to implement "one-click reverse image search" tools. Google's News Initiative offers such a tool, which empowers users to find factual information about an image—including its provenance, history of previous use, and any related images—through a series of simple steps.[117] Messaging platforms should consider collaborating with an Internet search function to enable recipients of images to quickly check the image for relevant information. Here again, WhatsApp leads the effort among messaging apps through its piloting of an "Internet search for forwarded messages" tool.[118] However, the tool is currently available in a minority of countries where WhatsApp operates.[119]

A third affordance that has been proposed, but mostly in theoretical discussions among scholars of content moderation in encrypted environments, is a hash matching and flagging system for known disinformation installed on users' devices. Such on-device matching and flagging is consistent with the privacy guarantees of end-to-end encryption *as long as the matching process and result are kept strictly on device*—that is, no one outside of the communication participants learns any new information.[120] This proposal is different from the so-called "client-side scanning" method, which *does* violate the privacy of users' communications because it is set up to automatically release information about a match to third parties, such as the platform or law enforcement agencies.[121]

The benefit of an on-device matching system is that it empowers users with information about content that they are about to send or receive, alerting them if it contains previously fact-checked information and informing them of the source of the fact-checking. Users retain full agency—they can decide to send or receive the message regardless of the flag. A similar "speed-bump" approach has been implemented on social media to caution users before sending images containing nudity—a warning that they can choose to ignore.[122]

Surveyed messaging app users seemed open to using such a tool: 89% of users across the nine countries said they would be interested in a scanning algorithm that informs them whether a piece of political content they receive on a messaging app is true or false according to specified fact-checkers. Furthermore, there is some evidence from social science research that content flags are effective in curbing the spread of misinformation.[123] Subject to further scientific and technical advancement, this tool may someday be able to alert users when they receive or are about to send synthetic content, such as deepfakes created with generative AI. However, platforms and fact-checking organizations should take care not to exacerbate the problem through their own ill-advised use of currently unreliable AI models to flag misinformation.[124]

The on-device matching method for known misinformation is still merely a viable proposal, but one that has not been tested or implemented. As such, platforms should consider supporting and monitoring academic research into such techniques but should not rush into rolling out a feature that might prove too intrusive or counterproductive.[125] A potential intermediate step might be to create optional plug-ins or extensions for on-device flagging that messaging app users can install on their devices. But doing so should not be a prerequisite for using the messaging service.

6 Choose whether to prioritize privacy or abuse mitigation—and be transparent with users.

When apps implement end-to-end encryption to protect users' privacy, they necessarily reduce their ability to monitor and combat abuse through direct content moderation. But the most widely used abuse mitigation approaches, even ones that are content-oblivious like metadata analysis and user reporting, also carry some risks to privacy. Platforms have to choose which values they wish to prioritize and be transparent with their user base about the implications.[126]

Some encrypted messaging platforms may choose to collect and analyze metadata—that is, information about a message, file, or user, rather than the content of the message or file itself[127]—in order to develop behavioral signals of abuse that help them detect suspicious behavior proactively. Following this approach, WhatsApp collects a profusion of metadata which allows the platform to combat spam-like activity, such as bulk and automated messaging.[128] Viber and Telegram also collect significant amounts of metadata, although their anti-spam systems seem to rely more on user feedback than on the metadata itself.[129] Signal, by contrast, collects the least amount of metadata necessary to provide its messaging service, even if doing so hampers its ability to detect abuse. And it does that for an important reason: Metadata collection carries significant risks for users' privacy.[130]

Whatever approach the platform takes, it should be forthcoming with its users. In this regard, WhatsApp is not transparent, stating in its marketing materials that "your privacy is our priority."[131] Yet, if privacy were truly WhatsApp's priority, it would refrain from collecting such large amounts of metadata. By contrast, Signal is honest in describing its product. In the words of its president, Meredith Whittaker, "we are not in the business of compromising on privacy"—and the app goes to great lengths *not* to collect any metadata, even if this means foregoing opportunities to mitigate abuse and to make money.[132]

Ultimately, even if a platform collects and analyses metadata, it can make efforts to minimize the risks associated with such collection and analysis. In the interest of its business, WhatsApp has chosen to monetize some of the metadata it collects by sharing it with its parent company, Meta, for targeted advertisements.[133] Such a business model is incompatible not only with data minimization best practices but also with the company's privacy assurances to its customers.

The same privacy considerations and transparency exhortations apply to user reporting—another commonly used platform moderation tool. User reporting is generally consistent with the privacy guarantees of end-to-end encryption, because platforms learn the content exchanged only after one of the parties to the communication chooses to share such content with the platform. Messaging app users overwhelmingly think that it is important to have the option to report a problematic user or content for the app to determine whether that user or content is appropriate.[134]

But, like metadata, user reporting carries privacy implications for users which they may not be aware of. First, the decision to report is made by only one of the parties to the communication, meaning that at least another of the parties usually has not consented to having their communications revealed to the platform. This is less of an issue when the party whose privacy is violated has committed an offense. But, in some instances, ill-intentioned actors have used the reporting system as a weapon and harassment tactic.[135]

Second, user reporting may leak more information to platforms than users realize. WhatsApp, for instance, collects the last five messages exchanged upon receiving a user report from a specific chat—a little known practice buried in WhatsApp's Frequently Asked Questions.[136] Whether platforms decide to enable user reporting or not, they should ensure the system is transparent, accessible, and effective. Doing so entails investing in a user-friendly interface, improving the quality and speed of report resolutions, and enacting safeguards against the exploitation of the reporting system itself.[137]

Finally, messaging platforms should invest more in user education regarding, for example, the level of privacy provided by different chat features and the type of platform moderation applied. WhatsApp deserves credit for its system of sending information and reminders to users via the official in-app WhatsApp channel. Other messaging platforms should follow this approach.

7 Create a clear separation between a platform's messaging functions and its broadcasting or networking functions, and consider bifurcating those aspects into separate products.

The tendency of messaging platforms toward feature bloat undercuts the platforms' mission as private and secure channels for communication and leads to confusion among users. WhatsApp, Telegram, and Viber—all of which have expanded into social media territory with their breadth of virality-promoting features—should establish a clear separation between their messaging service, meant for individual and small-group chats, and their social media features. In fact, the platforms should consider bifurcating their services into separate apps—one purely for private messaging, in which all content and metadata are protected with end-to-end encryption, and another for broadcasts, channels, stories, and large groups, in which content is left in plaintext and moderated rigorously. Establishing a clear separation between encrypted and non-encrypted functionalities would also help users have more accurate mental models of the security and privacy guarantees of each side of that separation, thereby minimizing the risks that come with having a false sense of security and privacy.[138]

Short of establishing such a bifurcation, platforms that continue to describe themselves as secure and private messaging services should seek to emulate Signal's anti-virality design (although even Signal has succumbed to user pressure to roll out a stories feature[139]). For example, they should add more friction to message forwarding, such as by eliminating simultaneous forwarding to more than five contacts at once. They should refrain from further increasing chat group sizes or rolling out supergroups, which start to approach social-media-style broadcasting. And they should employ reasonable group creation limits to filter out inauthentic activity.[140]

For policymakers

8 Include encrypted messaging platforms within the scope of online platform regulation, but ensure that compliance with the regulation does not entail breaking encryption.

The main recommendation for policymakers concerned about the exploitation of encrypted messaging by political propagandists is to not make the problem worse by imposing obligations on platforms that undermine encryption. Notwithstanding this report's focus on the issue of information manipulation, the value of encrypted messaging for human rights defenders, and society more broadly, exceeds the threat of disinformation on encrypted chat apps. Moreover, there are various approaches to combating information manipulation on these apps which do not require weakening encryption.

There is a trend among governments to demand that online platforms remove or weaken end-to-end encryption in the name of national security, child safety, or the need to quell disinformation.[141] Brazil,[142] India,[143] Indonesia,[144] and the UK,[145] have recently passed or proposed laws imposing client-side scanning, source tracing (also known as "traceability" mandates), or vague requirements to track and take down illegal content—all of which entail breaking encryption or sharply disincentivize its deployment.

Rather than passing regulations that undermine the use of encryption, policymakers should include encrypted messaging applications within the scope of regulatory interventions that focus on increasing transparency by technology companies. Requiring companies to improve access to their platforms by independent researchers and to disclose content-neutral information about their policies and enforcement systems, for example, would serve as catalysts for more effective self-regulation by encrypted messaging platforms.

9 Support effective bottom-up media literacy initiatives.

Policy makers should provide financial support and protection for media literacy initiatives run by non-partisan civil society groups with proven track records—particularly those that work with marginalized populations. Research shows that people are particularly likely to believe information—or change beliefs—based upon content received from people they know or care about.[146] But attempts to spread media-, digital-, and informational- literacy are often top-down and, as such, limited in their ability to impact the consumption of misleading or problematic information among particular social, cultural, and linguistic sub-groups.

Rather than helicoptering into particular communities and prescribing solutions developed absent community collaboration, policymakers should support "home-grown" or grassroots information resilience and literacy efforts.[147] A positive example comes from Brown University's Information Future's Lab, which partnered with We Are Más, a Florida-based micro-engagement agency, to work with trusted messengers in South Florida's diaspora communities to spread credible information.[148] Policymakers should meaningfully support and work alongside such pre-existing networks of community groups and their media literacy ventures.[149]

For civil society and researchers

10 Contribute toward media literacy efforts.

The success of tiplines and other media literacy efforts depends on the participation of credible news and civil society organizations. Rather than relying solely on internet traffic, these organizations should meet messaging app users where they are—in the apps themselves. In addition to running tiplines, they can use channel and broadcast features to promote news from credible sources, run misinformation "prebunking" campaigns,[150] educate users on ways to identify manipulated and synthetic content, and raise awareness about in-app behaviors that exacerbate the spread of misinformation.[151]

The International Fact-Checking Network (IFCN) can play a catalyzing role by recruiting more publications with diverse audiences to join its network. As Michael Rain, the founder of a media and research company, points out, the network should enlist publications that serve underrepresented and immigrant communities, such as Univision's El Detector fact-checker, "so that WhatsApp users will be more likely to utilize a fact-checking service that is culturally relevant to them."[152]

Researchers can also support these efforts by undertaking studies on the effectiveness of different media literacy efforts, with a view to improving and scaling the most successful methods as well as understanding their limitations.[153]

11 Develop ethical methodologies for studying encrypted messaging platforms.

Disinformation on encrypted messaging platforms is especially challenging to study because of content encryption and users' expectations of privacy on those platforms. Yet, as this report has shown, information that circulates on encrypted chat apps has implications for democracy and the public's access to reliable information.

Rather than being discouraged by these challenges, researchers should develop innovative methods to examine phenomena in encrypted settings. Professor Kiran Garimella's research group, which is devising ethical methods to access and analyze information circulating on WhatsApp,[154] provides an example to follow. The user survey conducted as part of this report is another method to build upon, and researchers are encouraged to use the survey results (contained in Appendix II) to derive additional findings and formulate new avenues for research.

Acknowledgements

This study is a joint project of the NYU Stern Center for Business and Human Rights and the Center for Media Engagement at The University of Texas at Austin. Research for this project was supported by Peter A. Horvitz, Clifford Ross, the Open Society Foundations, the Omidyar Network, the John S. and James L. Knight Foundation, and The Miami Foundation. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding bodies.

This project benefited from the research assistance of Riya Guha, Mihir Chhatre, Gabrielle Beacken, Zelly Martin, Tanvi Prem, and Meera Hatangadi. We would also like to thank the interviewees for sharing their time and insights.

Appendix I – Research Methodology

Qualitative data

The Propaganda Research Lab at UT Austin conducted 92 semi-structured interviews in person or online in 2023 and 2024 in Bolivia, Hungary, India, Mexico, Nigeria, Turkey, Tunisia, and the U.S. Previous interviews conducted in 2021-2023 in 17 countries provided background data for this project.

The interviewees included political propagandists and former political propagandists who have participated in the dissemination of disinformation on messaging apps; civil society activists, including individuals who organize for democracy, fact-checkers, and employees of non-profit organizations working to mitigate false and misleading information; trackers of political communication such as journalists and open-source intelligence analysts; and academics, who assisted in triangulating the insights from these diverse groups.

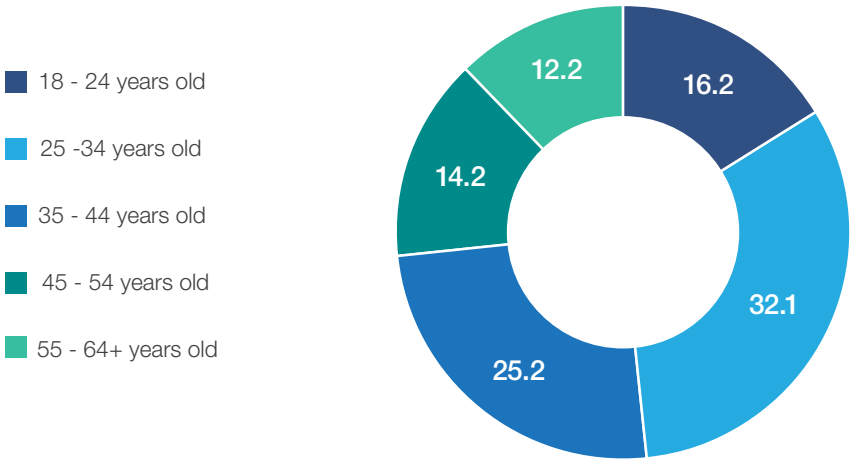
UT Austin's Lab contacted members of civil society and researchers by email, and then leveraged snowball sampling from those interviews to reach the hidden or hard-to-reach populations of activists and political propagandists.[155] The Lab further relied on its networks of hidden populations over the last seven years to contact potential interviewees. After conducting interviews, UT Austin's Lab engaged in a thematic coding process in which researchers collapsed codes from most specific to broadest, until they arrived at cross-cutting themes.[156]

Survey data

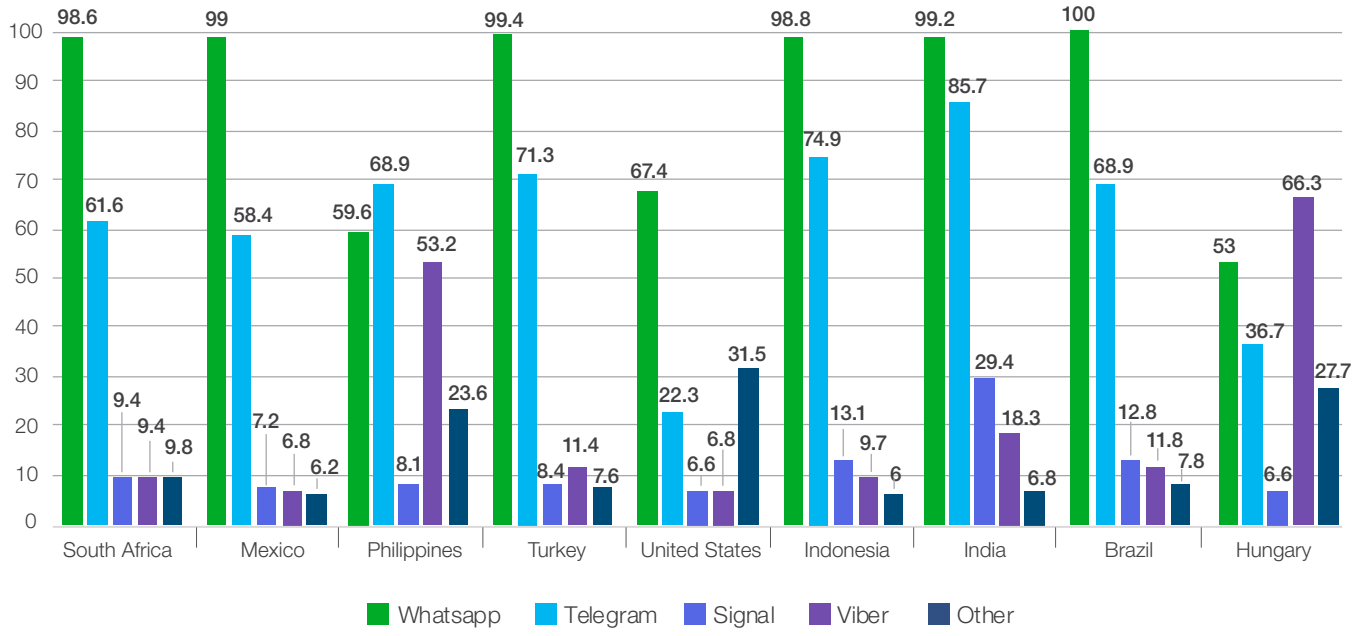
The survey of messaging app users was written and commissioned by the NYU Stern Center for Business and Human Rights, with financial support from Peter A. Horvitz. The survey questionnaire consisted of 29 questions (the full questionnaire can be found in Appendix II). The survey programming, translation, and data collection in nine countries was carried out by Dynata, an international data-collection company,[157] and overseen by NYU Stern, in May 2024. The nine countries covered were: South Africa, Mexico, the Philippines, Turkey, the U.S., Indonesia, India, Brazil, and Hungary. In total, 4,586 people completed the survey. Anonymized raw data with survey responses is available upon request.

Appendix II – Survey Results

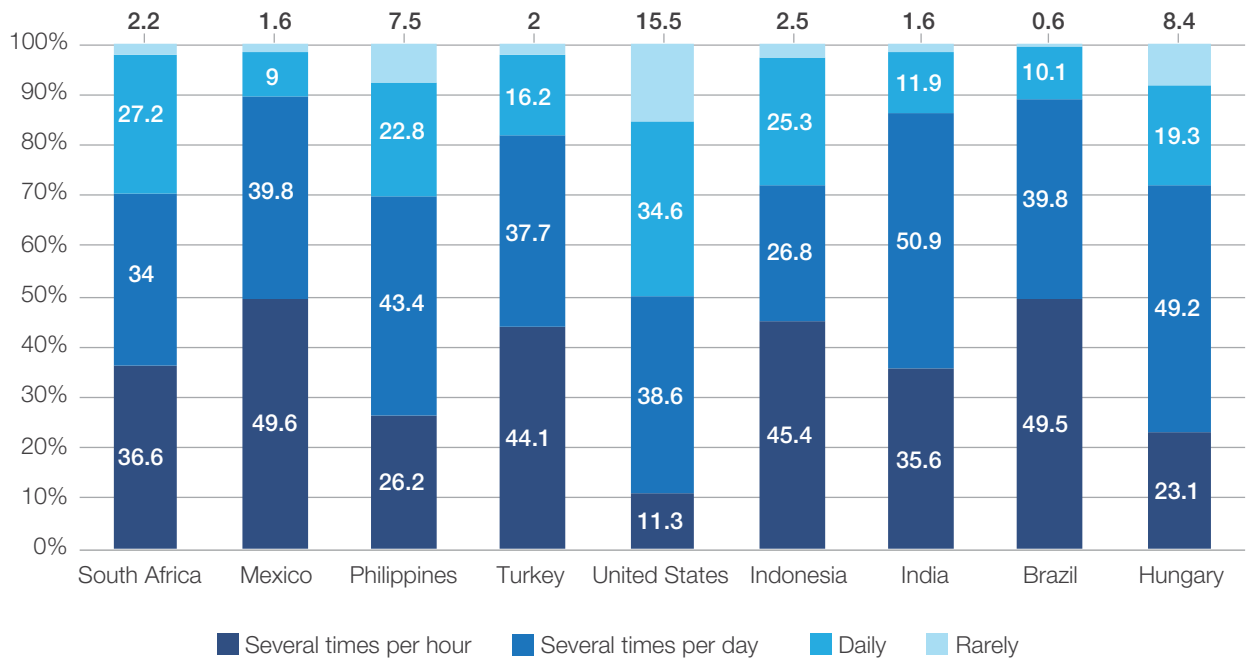
Q1. How old are you?



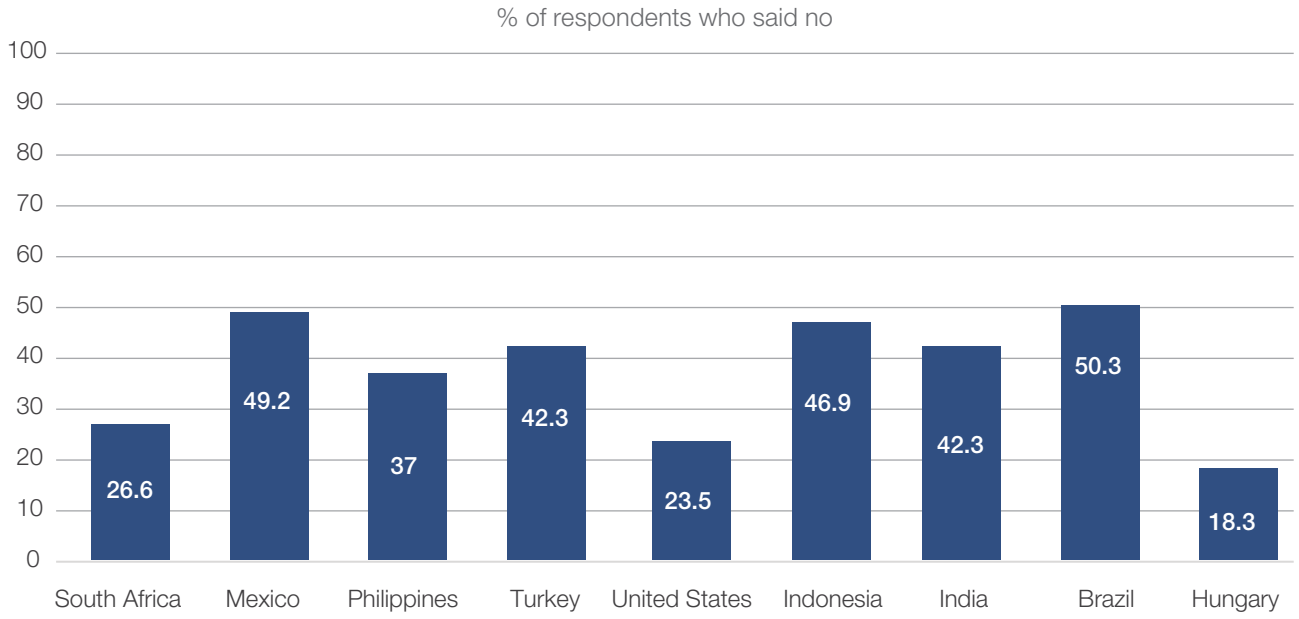
Q2. In the last year, have you used any of the following messaging apps?



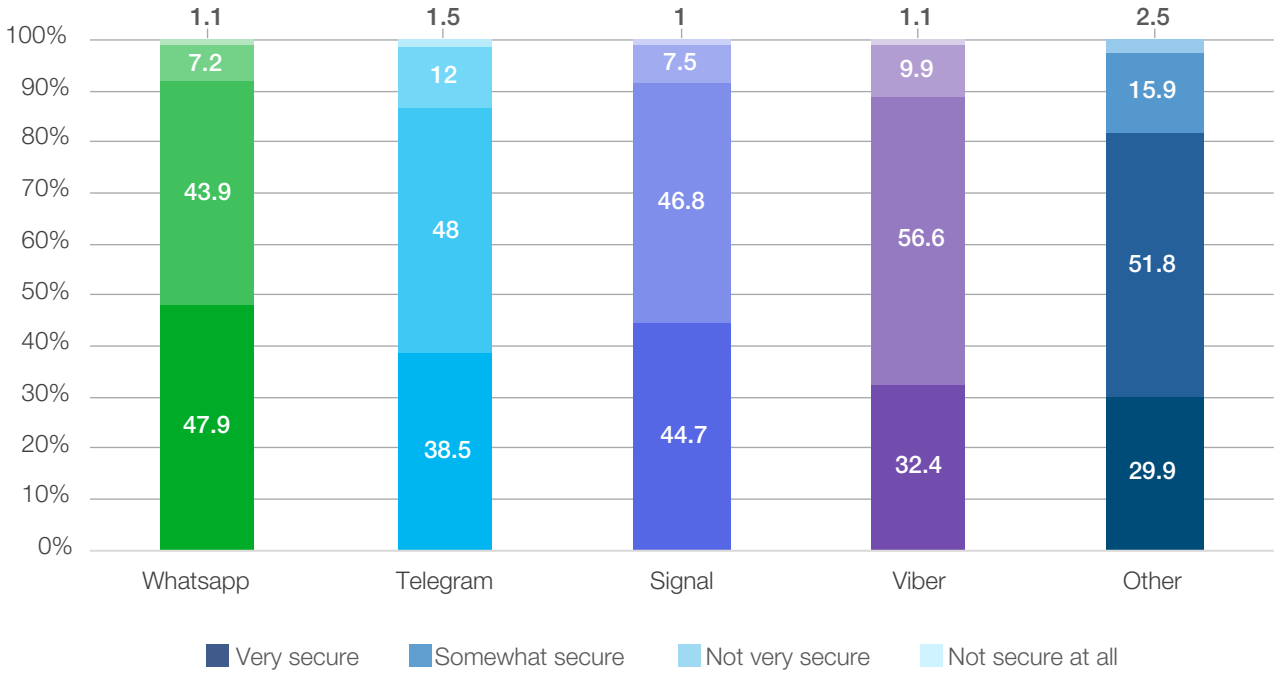
Q3. How frequently do you use one or more messaging apps?



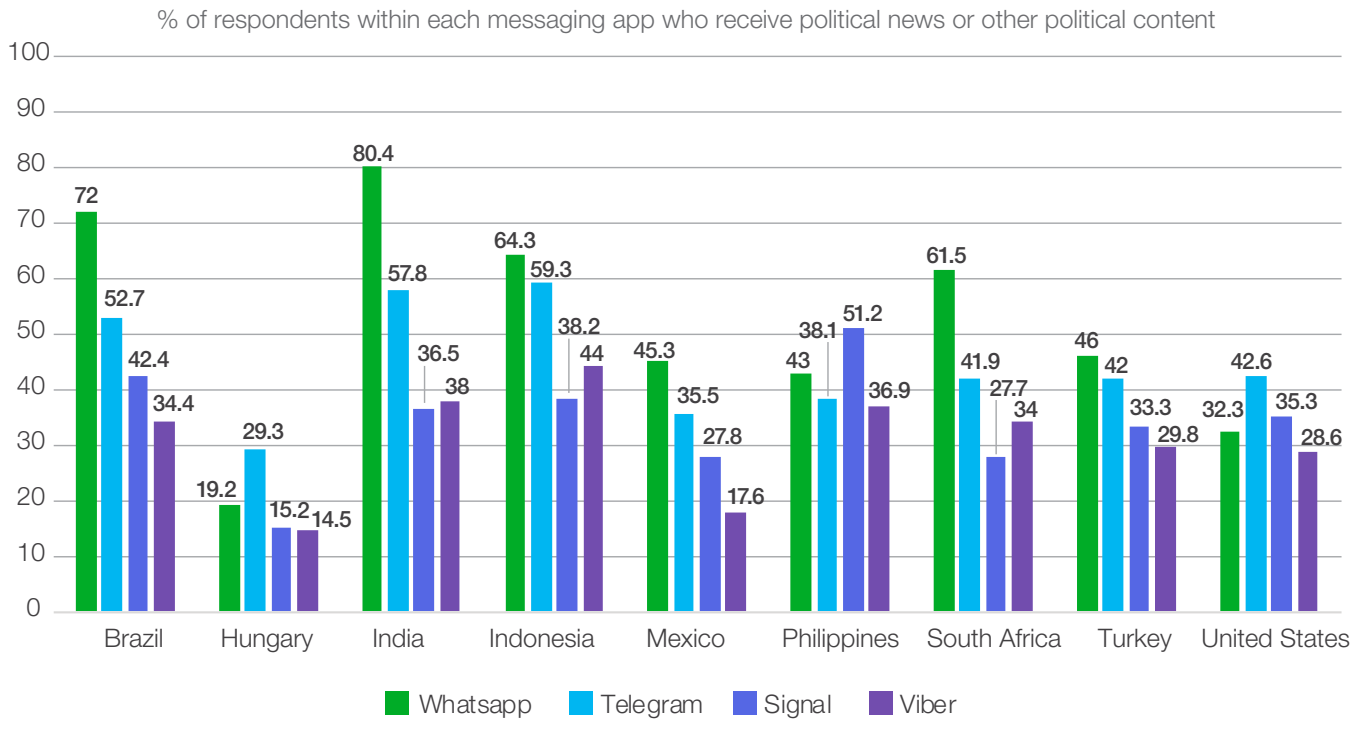
Q4: Could you delete messaging apps from your phone and keep up communication with friends and family via other means?



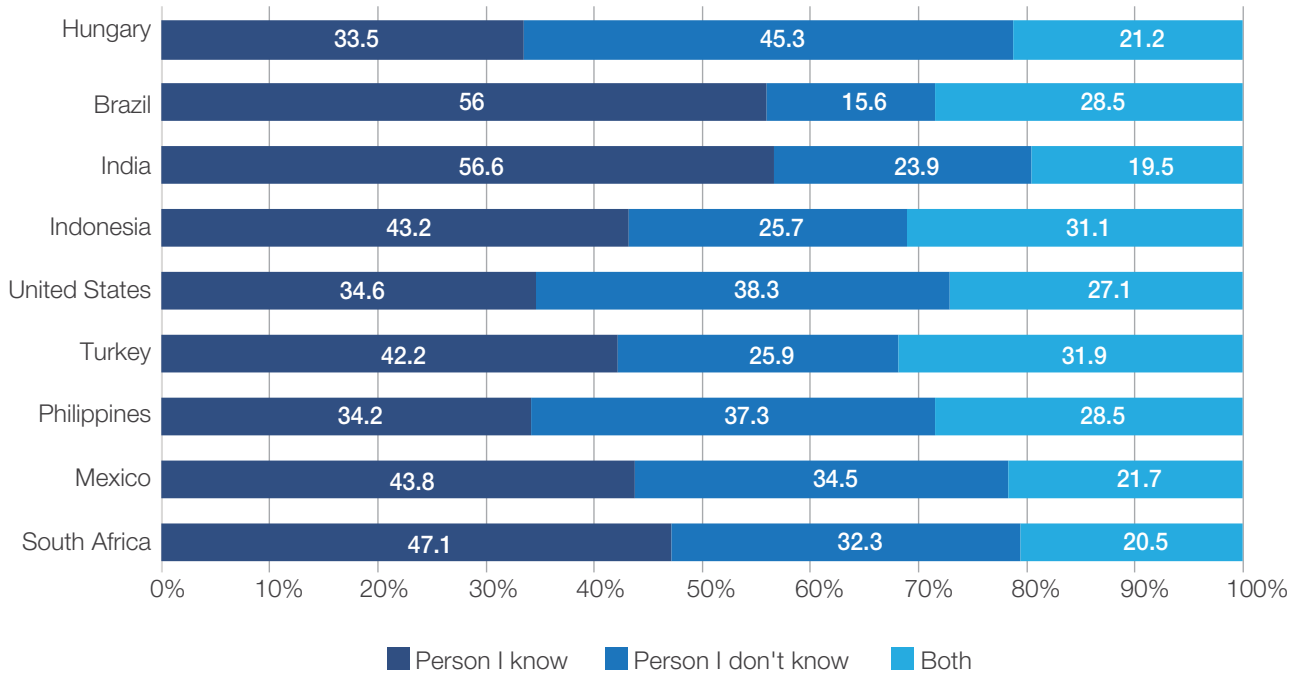
Q5: How secure and private do you think your messages are on messaging apps?



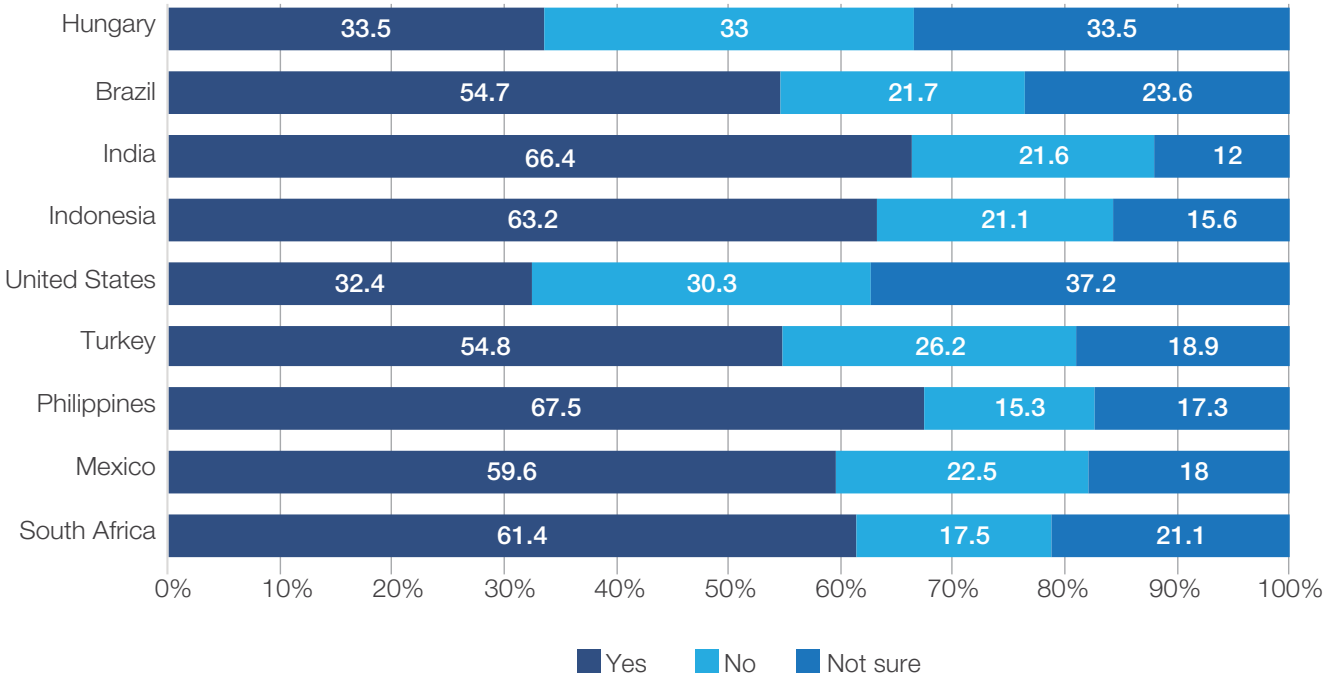
Q6: In the last year, have you received political news or other political content via one or more messaging apps?



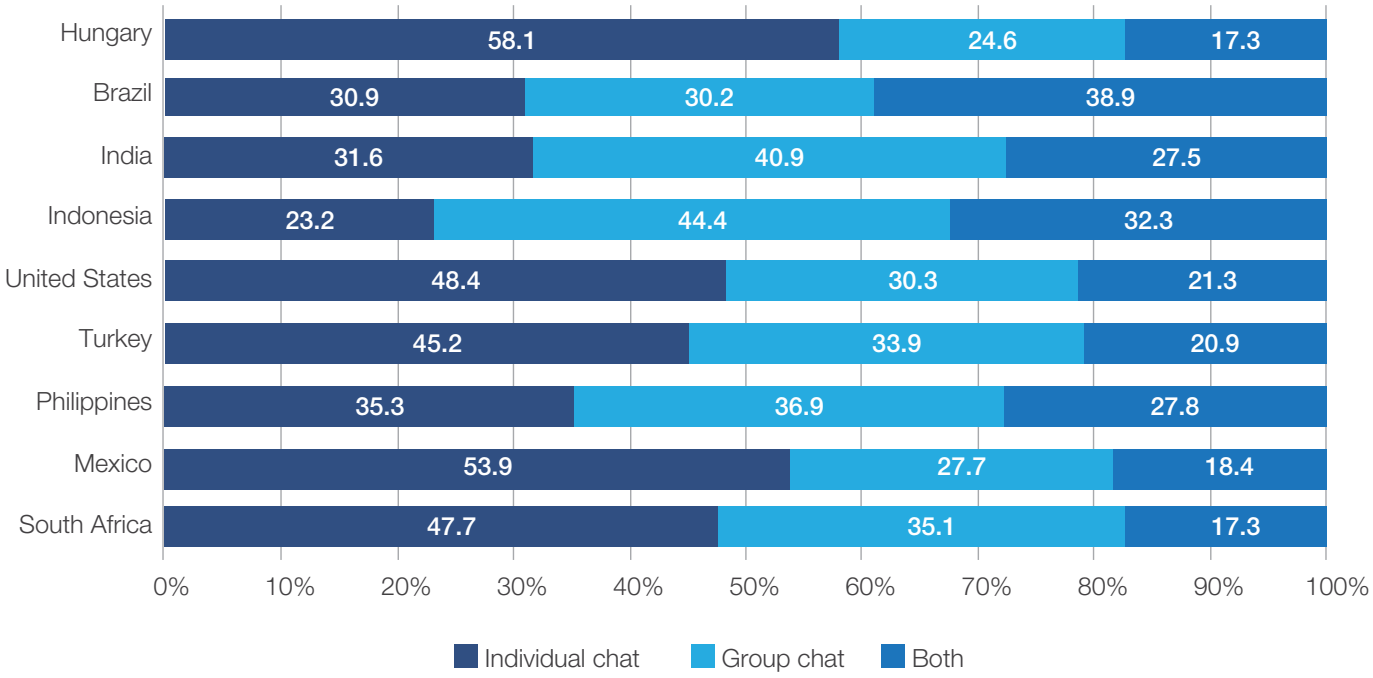
Q7: Did the political content come directly from a person/account you know, a person/account you don't know, or both?



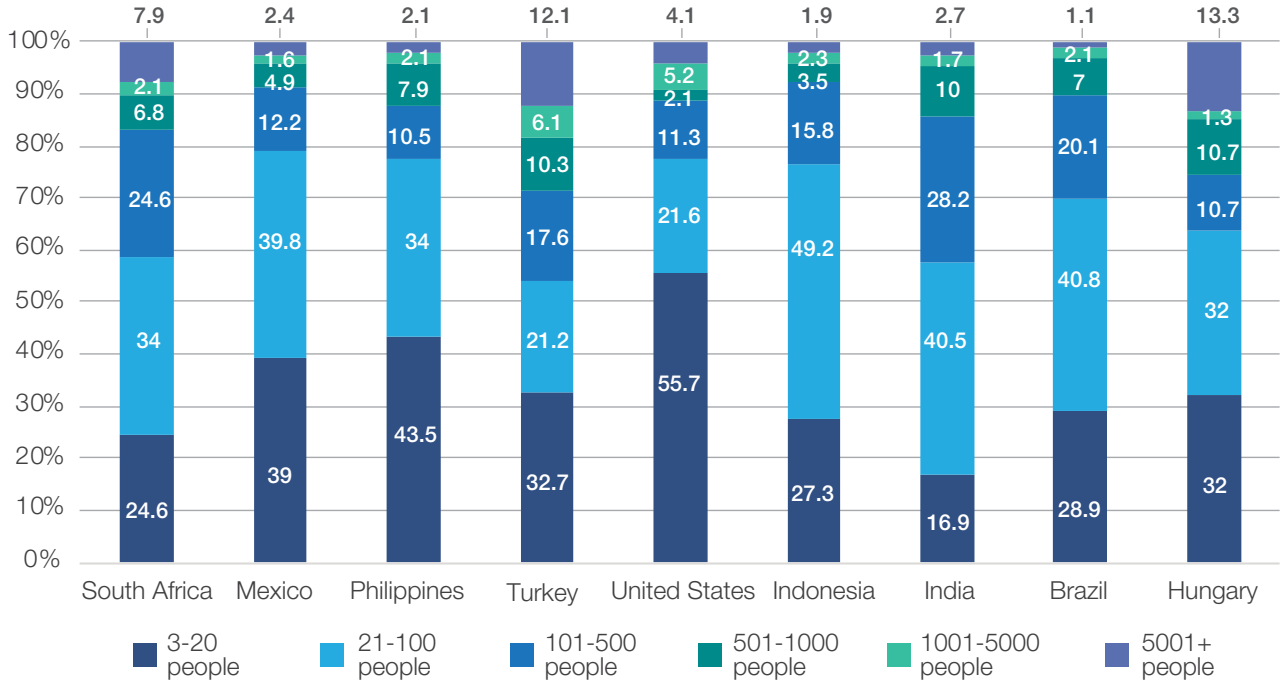
Q8. Was this political content forwarded from another chat?



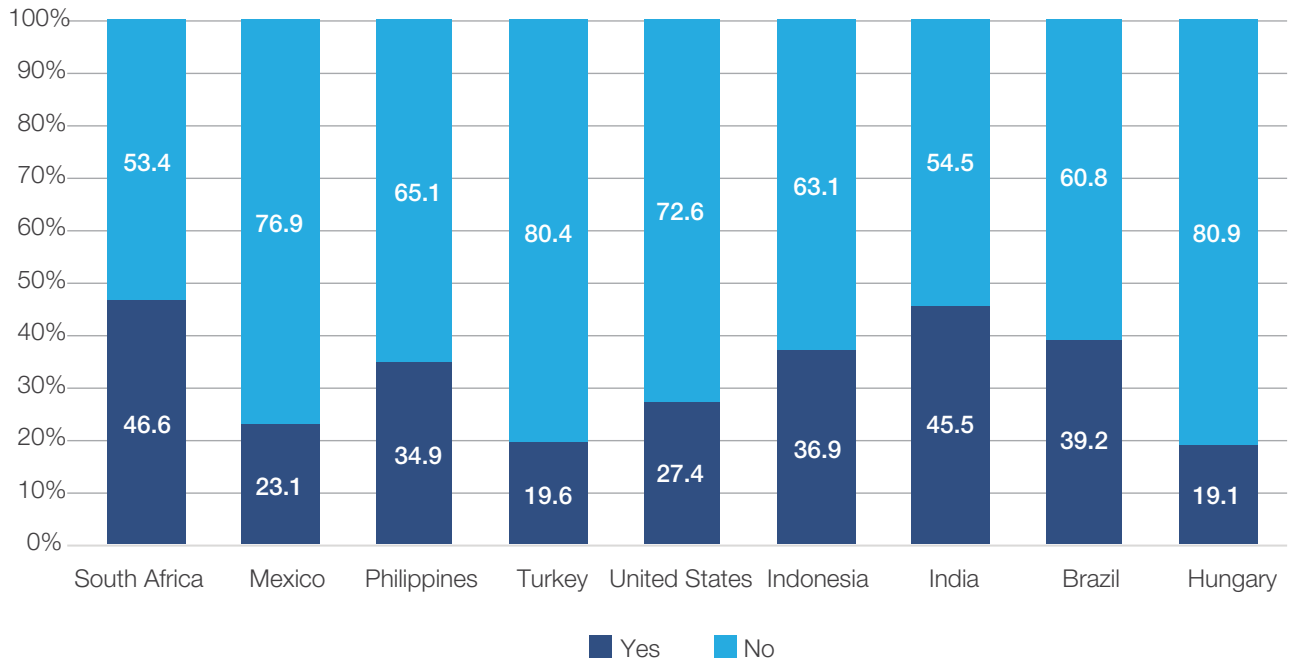
Q9: Did you receive the political content in an individual one-on-one chat or in a group chat?



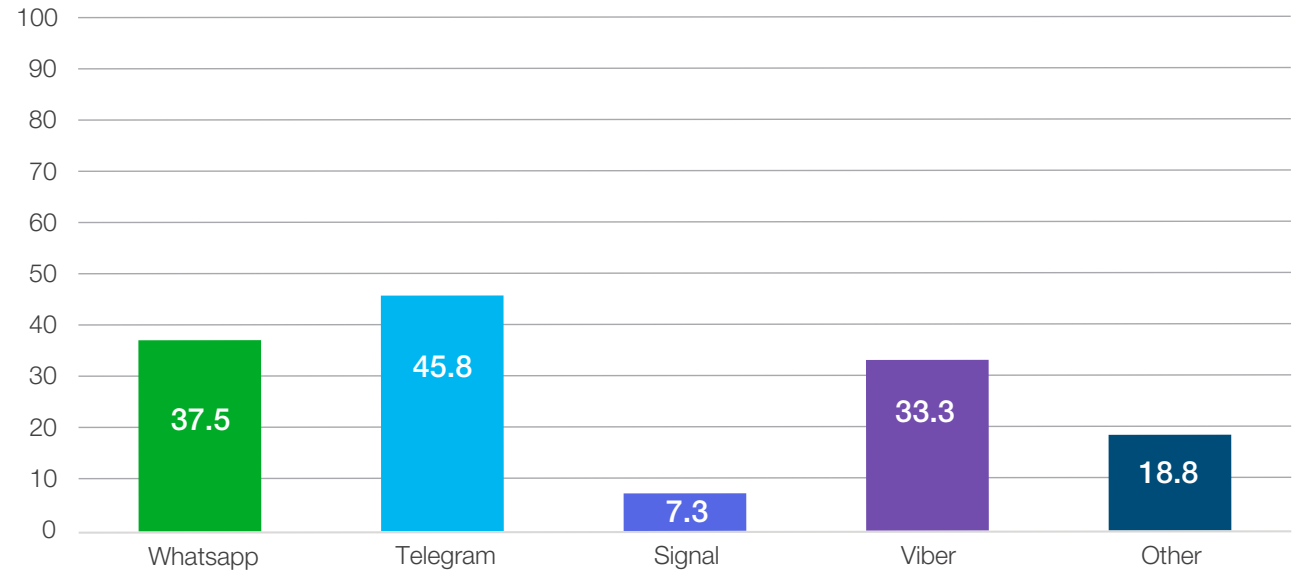
Q10: How many people were in the group in which the political content was shared?



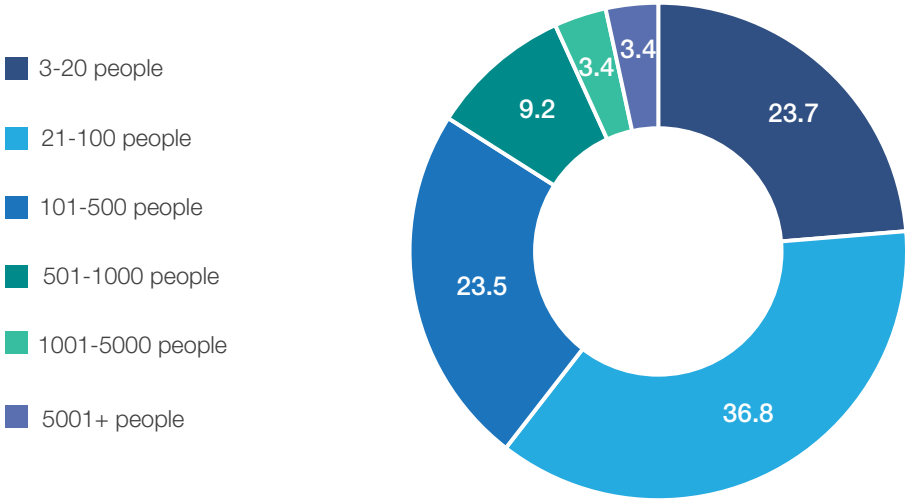
Q11: Have you been added, without your consent, to groups on a messaging app where political content is shared?



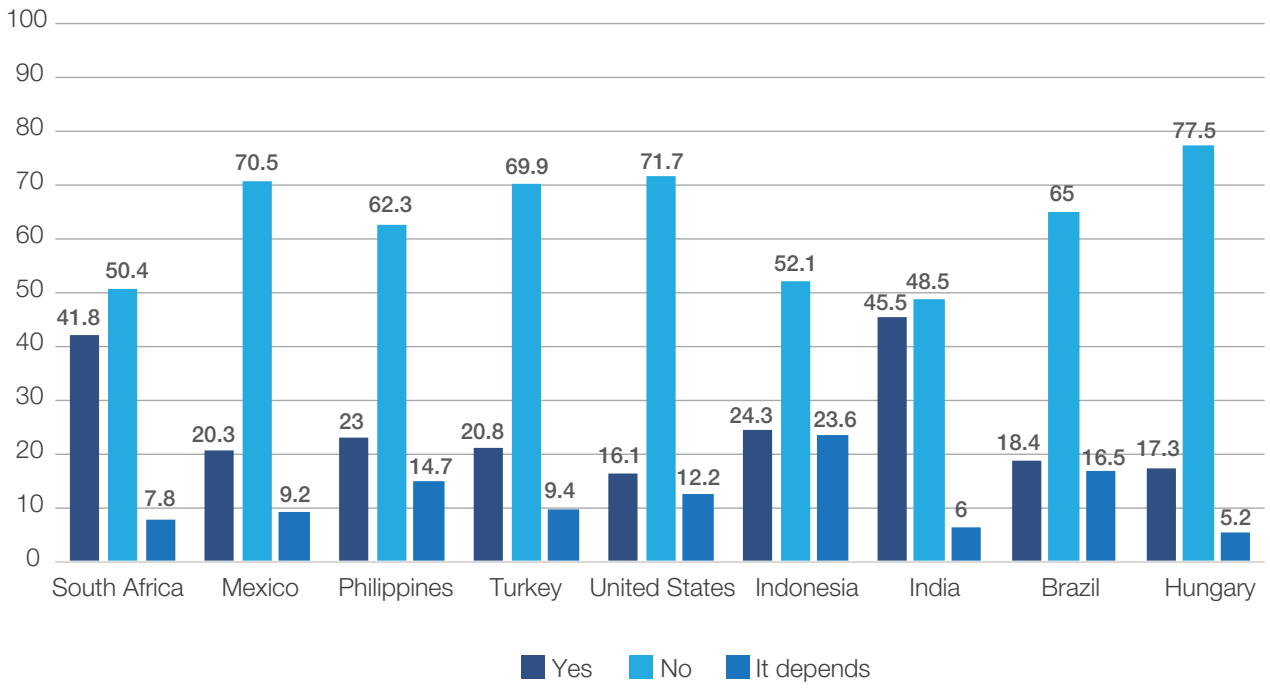
Q12: In which messaging app were you added to such groups?



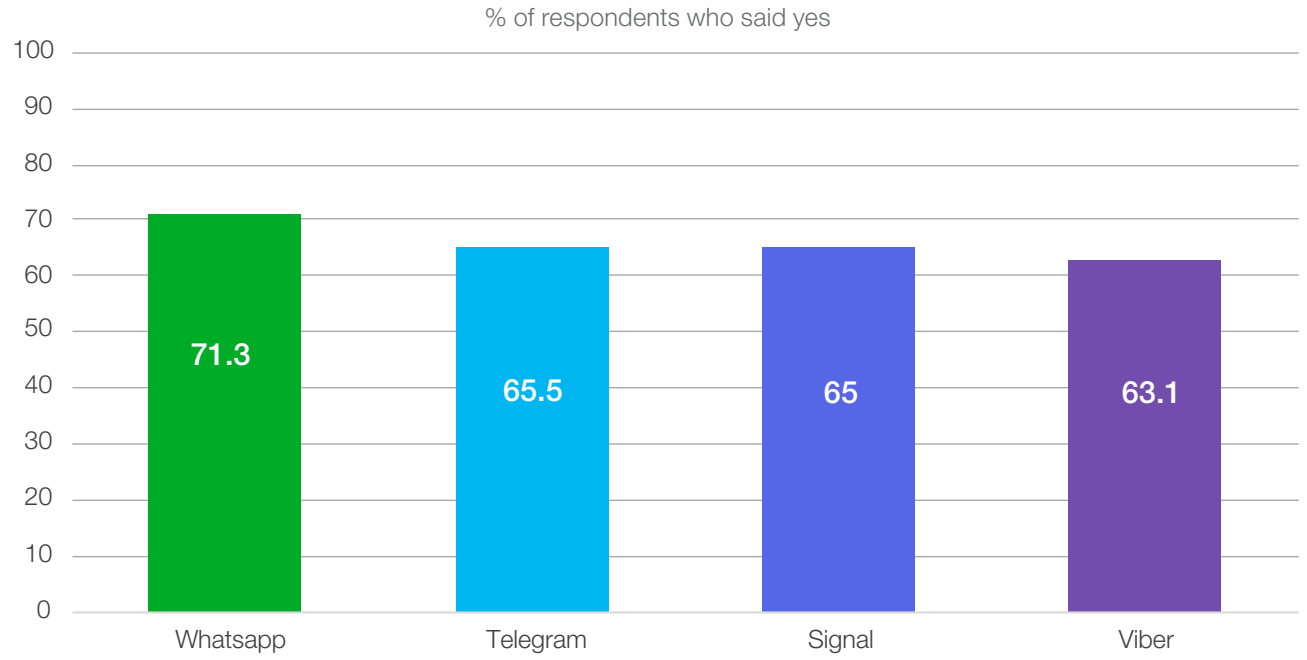
Q13: How large were these groups?



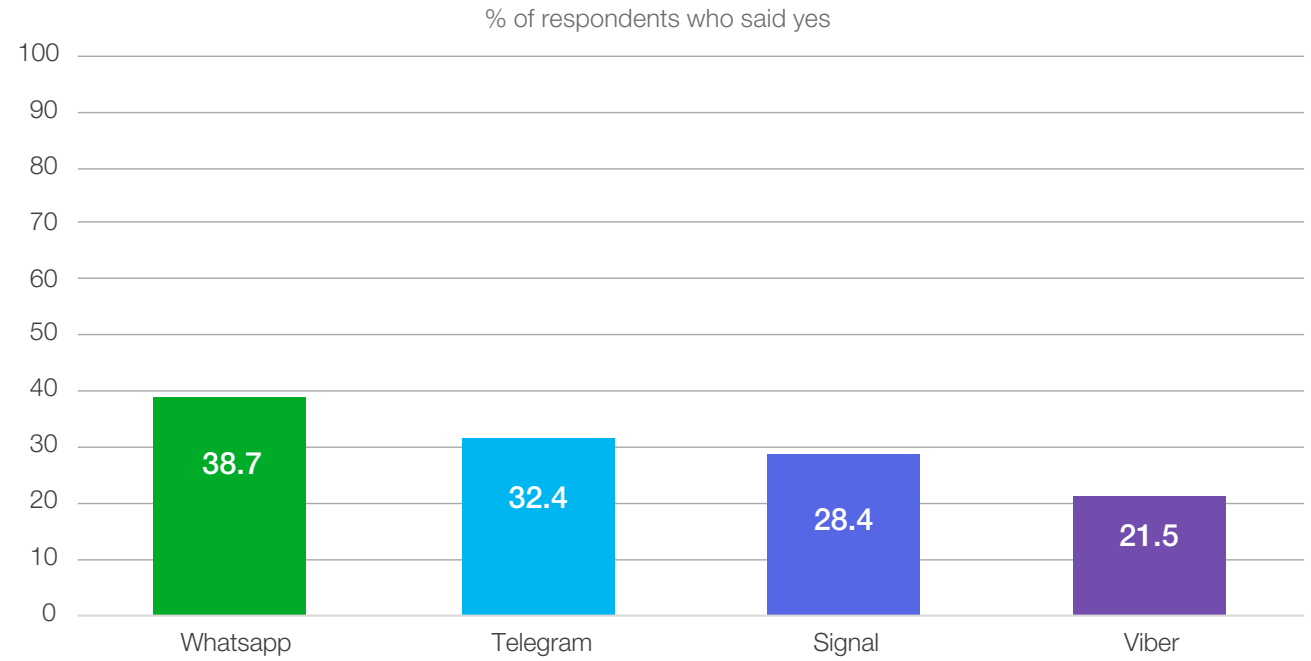
Q14: Do you trust the political content you receive on messaging apps?



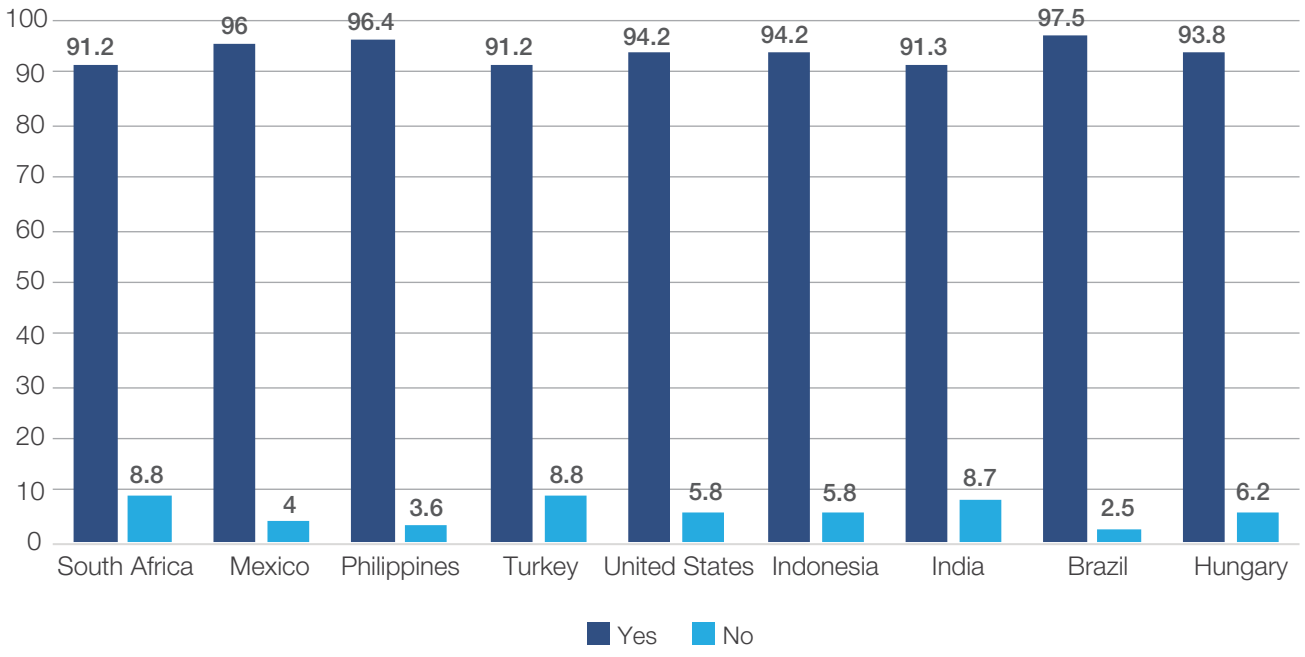
Q15: Are you aware of the option to report another user or content on [each messaging app previously selected] for the company to determine whether that user or content is appropriate?



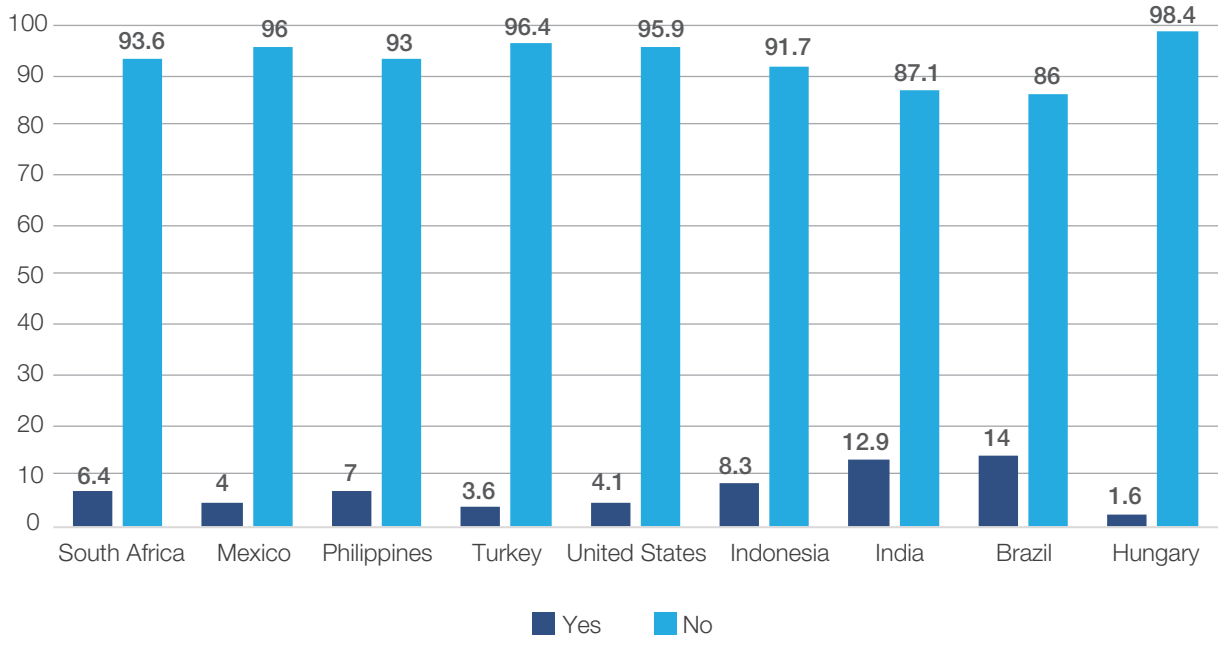
Q16: Have you ever reported another user or content?



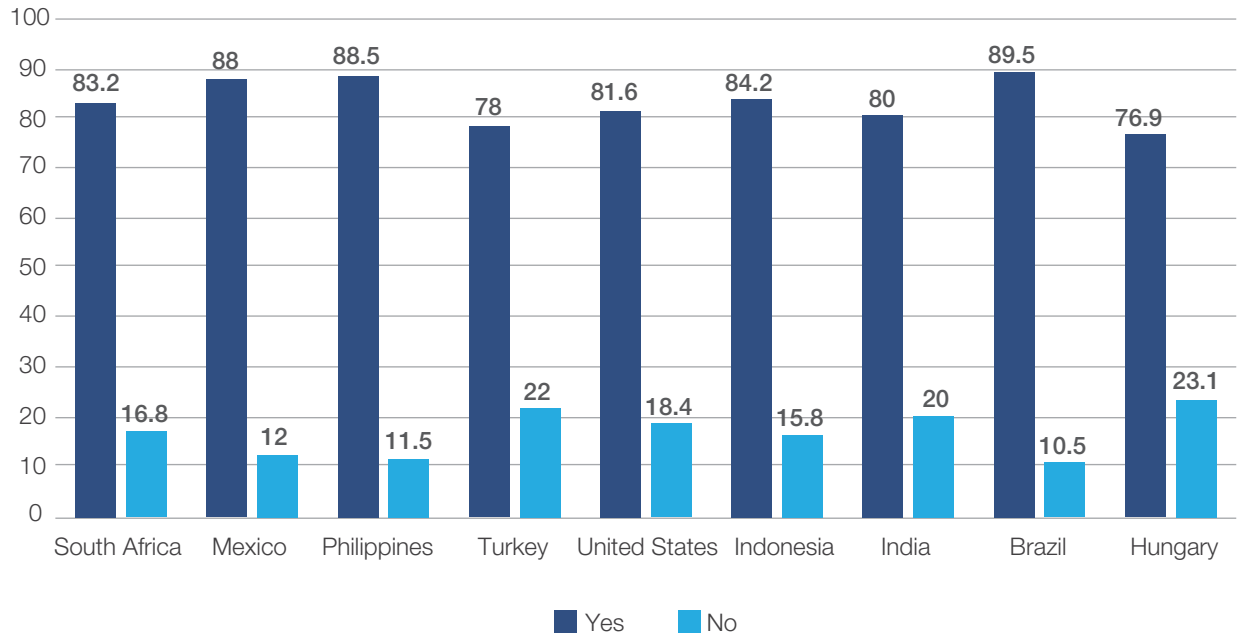
Q17: Do you think it is important to have the option to report a problematic user or content to the messaging app for the app to determine whether that user or content is appropriate?



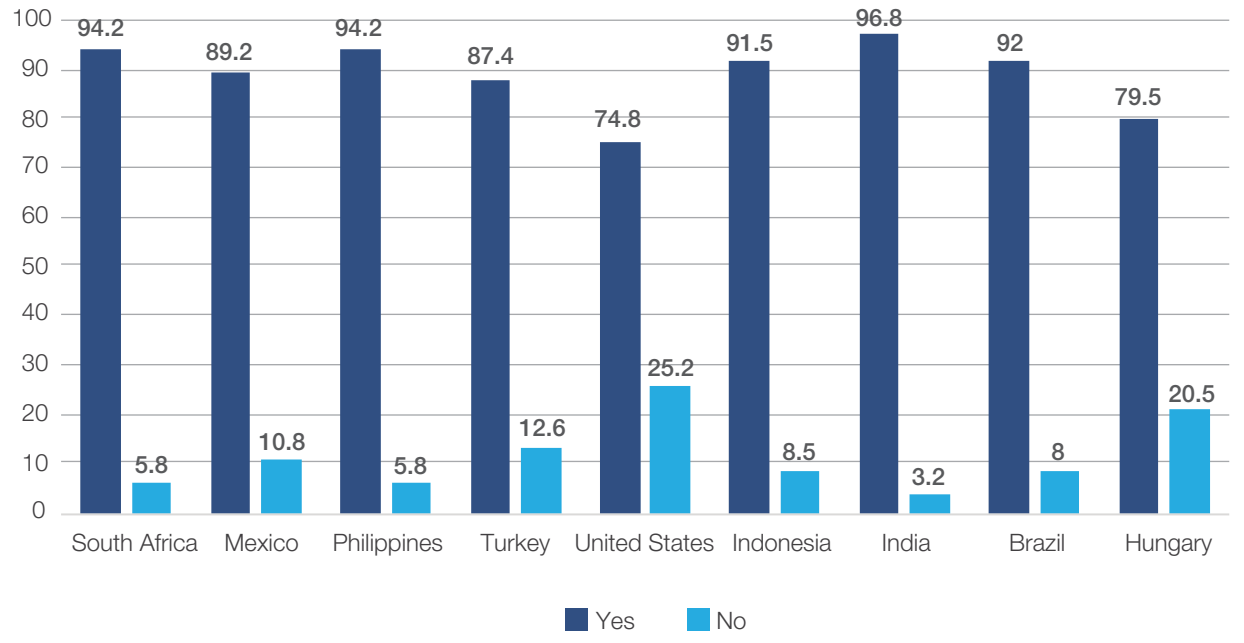
Q18: Have you ever contacted a "tipline" or fact-checking organization to check the truthfulness of a piece of political content received on a messaging app?



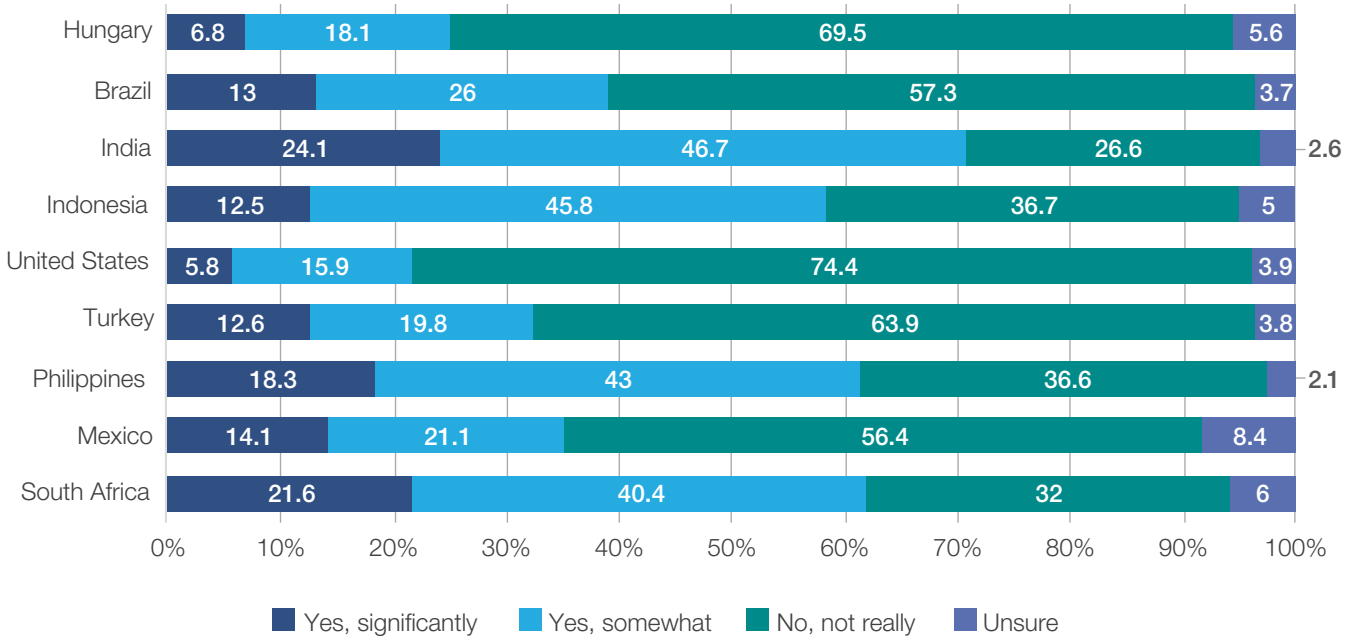
Q19: Do you think it would be useful to be able to contact a "tipline" or fact-checking organization to check the truthfulness of a piece of political content received on a messaging app?



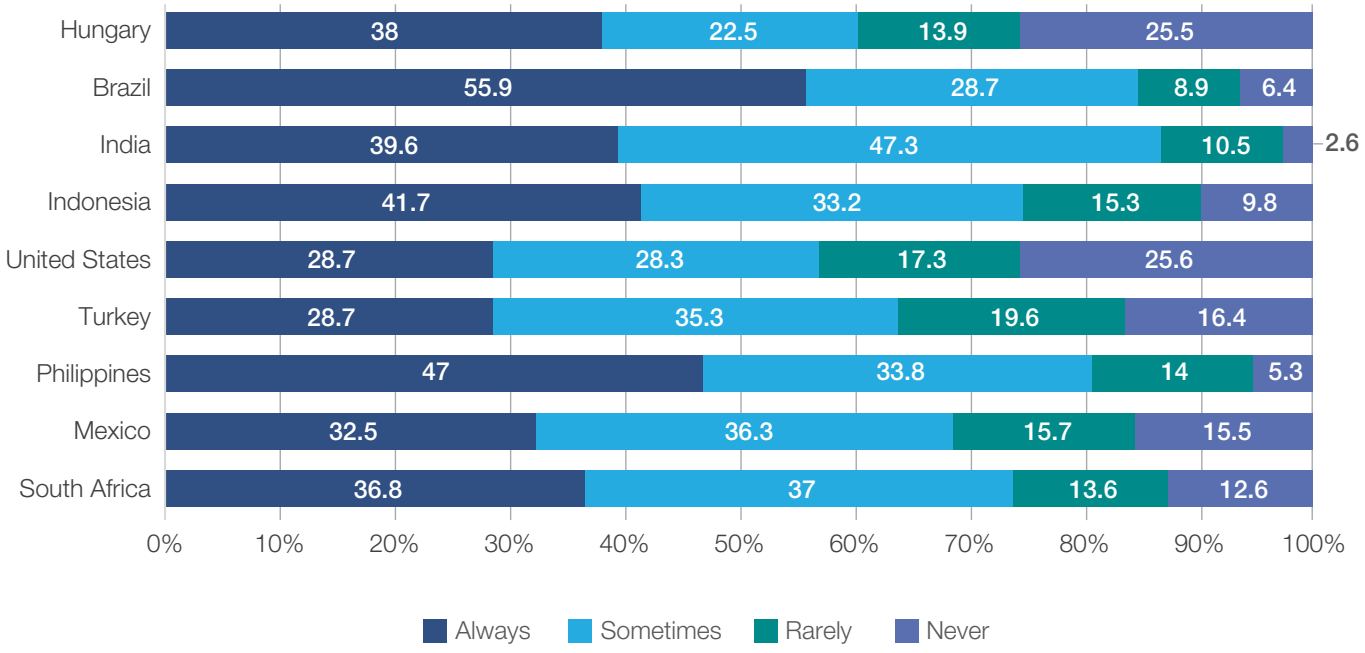
Q20: Would you be interested in a scanning algorithm that informs you whether a piece of political content you receive on a messaging app is true or false?



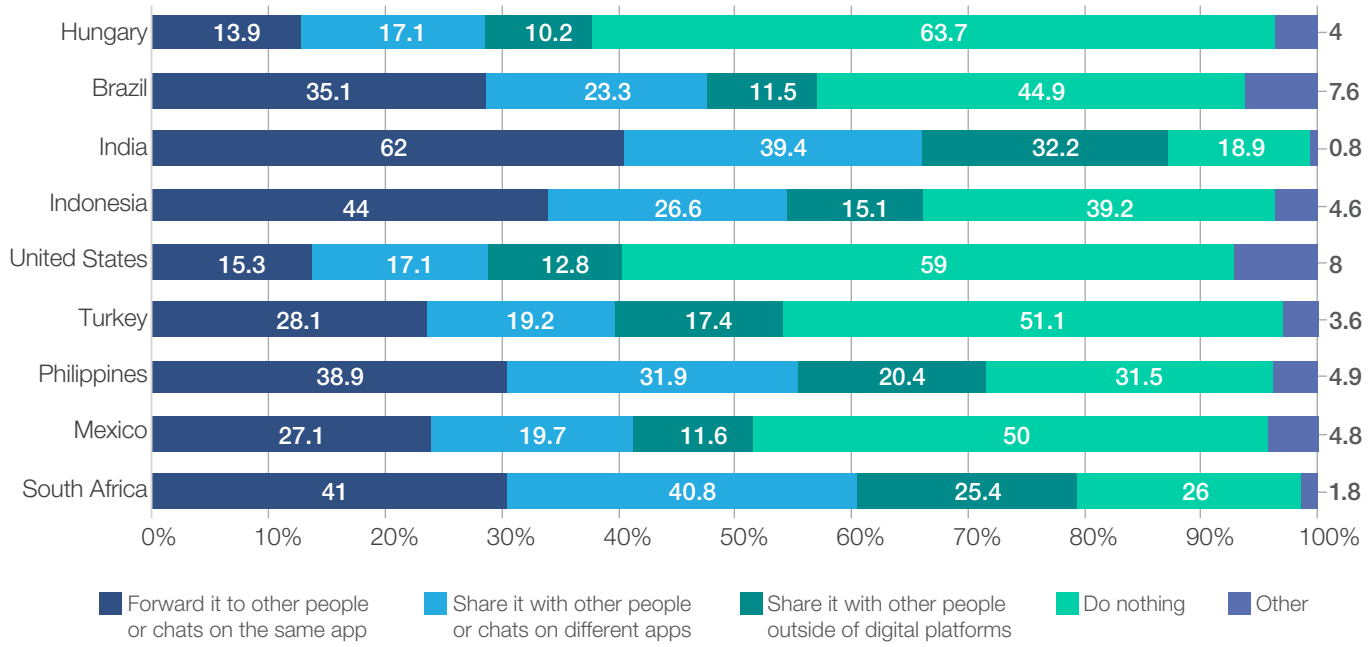
Q21: Do you believe political content shared on messaging apps has influenced your political opinions or beliefs?



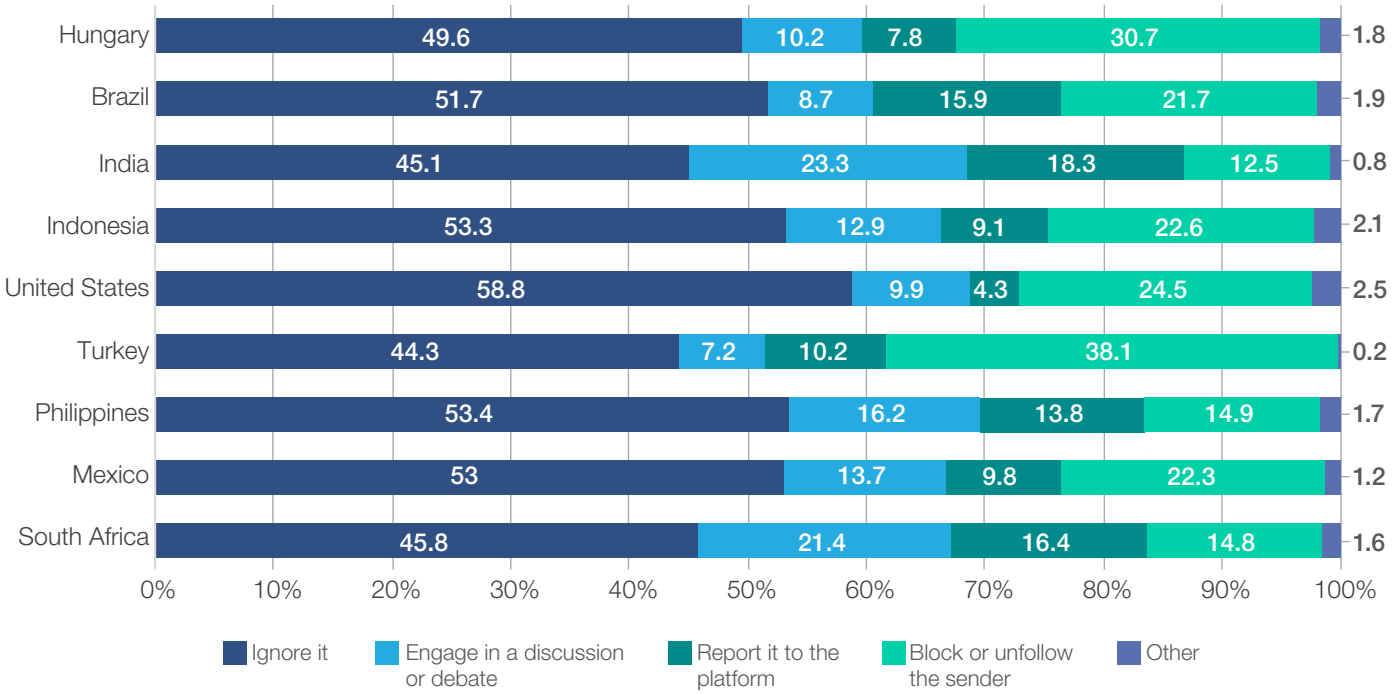
Q22: How often do you fact-check political content before sharing it on messaging apps?



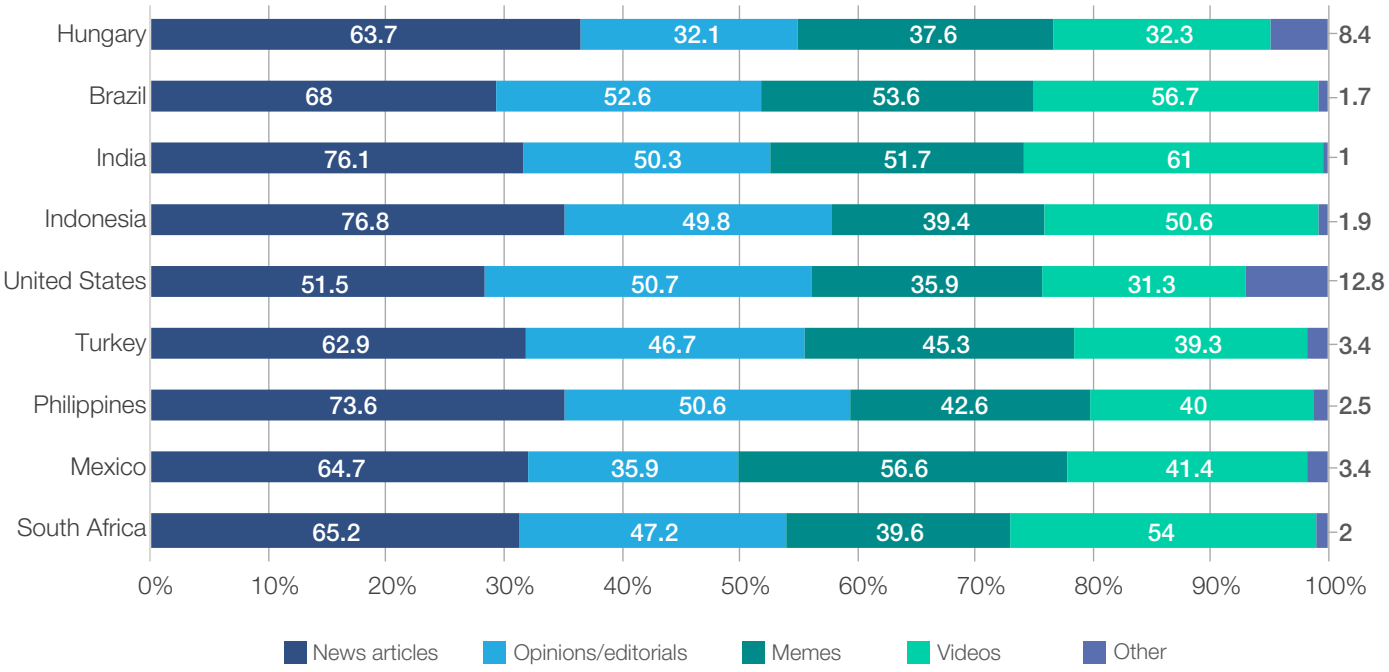
Q23: How do you typically respond when you encounter political content on messaging apps that you agree with or find interesting?



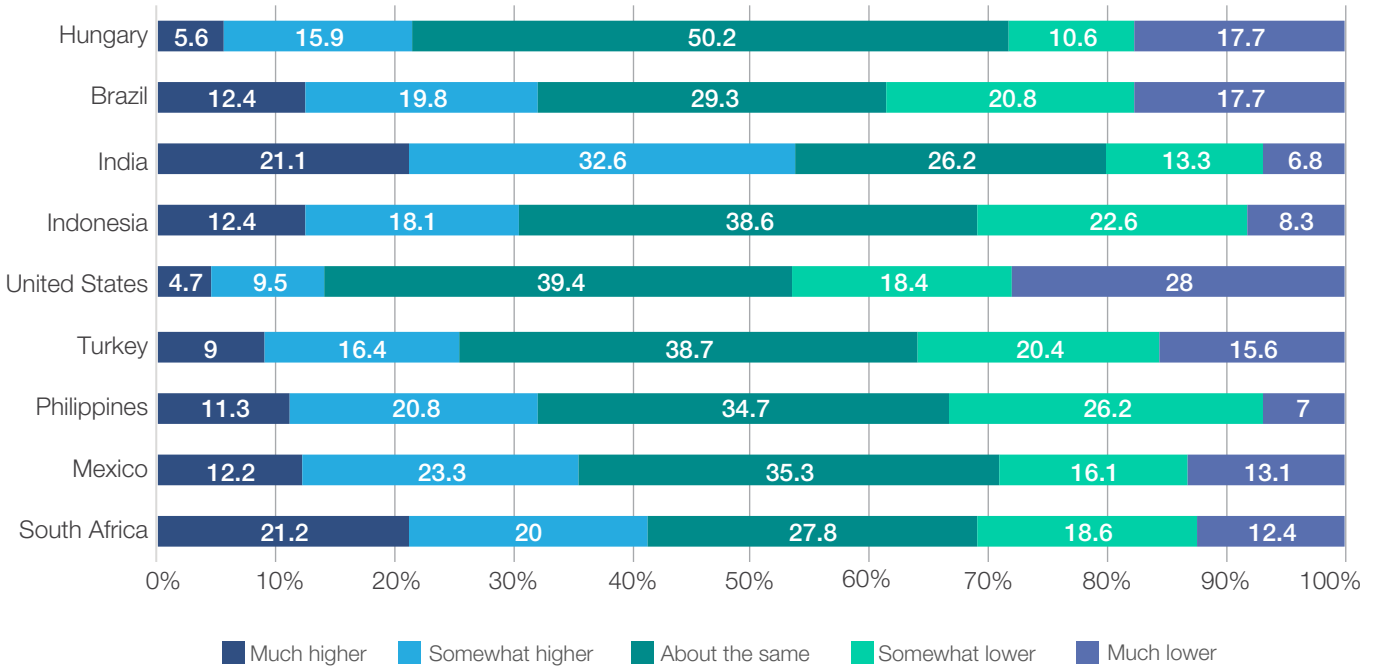
Q24: How do you typically respond when you encounter political content on messaging apps that you disagree with or find offensive?



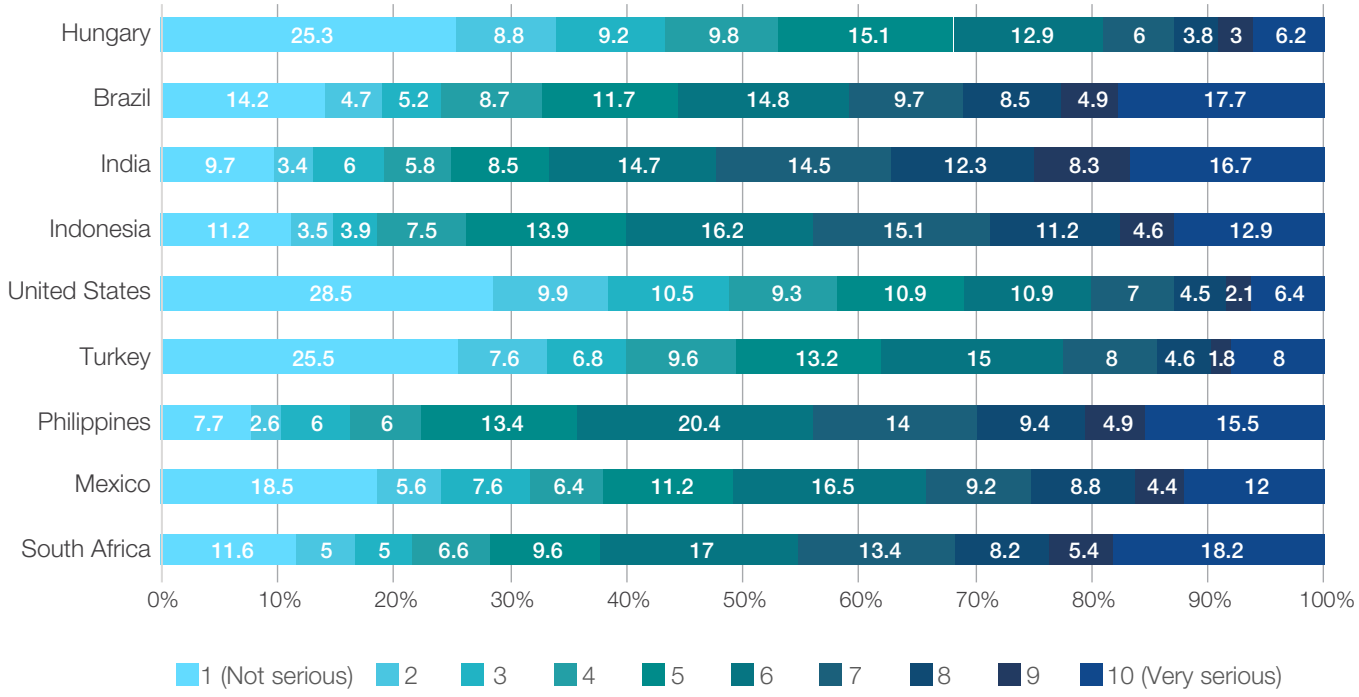
Q25: What types of political content do you typically encounter on messaging apps?



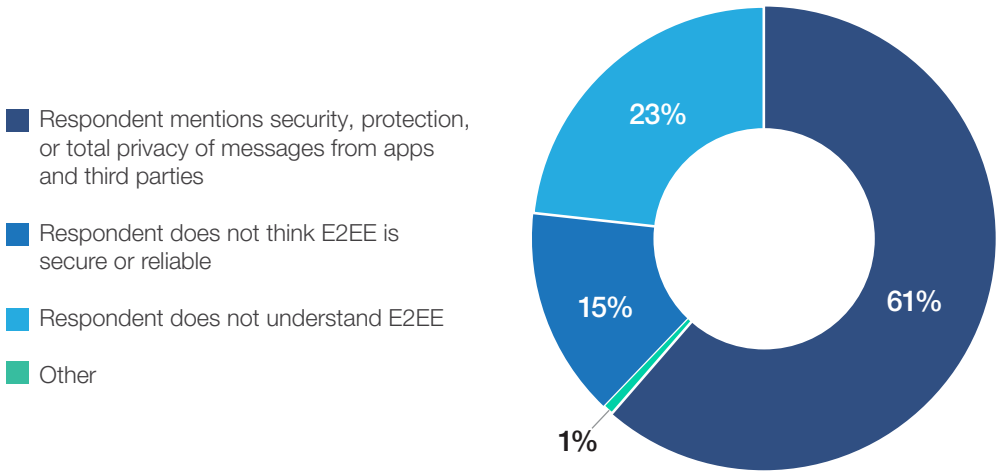
Q26: How would you rate your overall trust in political content shared on messaging apps compared to other news sources (such as TV, newspapers, Facebook, etc.)?



Q27: On a scale of 1 to 10, how seriously do you take political messages shared on messaging platforms, with 1 being not serious at all and 10 being very serious?



Q28: How do you understand the concept of end-to-end encryption (E2EE) as applied to messaging apps?*



**Question called for an open-ended response. Responses were coded and aggregated to create the following categories for visual representation.*

Endnotes

1 https://www.reuters.com/world/europe/telegram-messaging-app-ceo-pavel-durov-arrested-france-tf1-tv-says-2024-08-24/

2 See, e.g., https://www.techpolicy.press/the-arrest-of-telegrams-pavel-durov-whats-encryption-got-to-do-with-it/

3 https://www.usatoday.com/story/news/investigations/2024/08/29/pavel-durov-telegram-arrest-extremist-users/74990169007/

4 https://www.washingtonpost.com/technology/2024/08/29/telegram-app-durov-free-speech-child-sexual-abuse-images/

5 https://foreignpolicy.com/2021/03/13/telegram-signal-apps-right-wing-extremism-islamic-state-terrorism-violence-europol-encrypted/

6 Chaturvedi, S. (2016). *I am a troll: Inside the secret world of the BJP's Digital Army. Juggernaut.*

7 https://restofworld.org/2024/bjp-whatsapp-modi/; https://www.ft.com/content/9fe88fba-6c0d-11e9-a9a5-351eeaef6d84

8 See appendix II (survey) question 2.

9 https://www.indiatoday.in/technology/features/story/whatsapp-is-now-10-years-old-and-here-is-a-look-at-how-it-grew-changed-the-world-1465208-2019-02-26

10 Ozawa, J. V., Woolley, S. C., Straubhaar, J., Riedl, M. J., Joseff, K., & Gursky, J. (2023). How disinformation on WhatsApp went from campaign weapon to governmental propaganda in Brazil. *Social Media + Society*, 9(1), 1-11. https://doi.org/10.1177/20563051231160632

11 Trauthig, I. K., Martin, Z. C., & Woolley, S. C. (2024). Messaging Apps: A Rising Tool for Informational Autocrats. *Political Research Quarterly*, 77(1), 17-29. https://doi.org/10.1177/10659129231190932

12 See appendix II (survey) questions 3 and 4. https://www.businessofapps.com/data/messaging-app-market/

13 Center for Democracy & Technology (2021). Outside Looking In: Approaches to Content Moderation in End-to-End Encrypted Systems. *Center for Democracy & Technology*. https://cdt.org/insights/outside-looking-in-approaches-to-content-moderation-in-end-to-end-encrypted-systems/; see also https://www.internetsociety.org/issues/encryption/what-is/.

14 https://www.internetsociety.org/resources/doc/2021/factsheet-how-encryption-can-protect-advocacy-groups-and-social-change-movements/

15 See, e.g., https://www.nytimes.com/2020/06/11/style/signal-messaging-app-encryption-protests.html; https://www.wsj.com/articles/whatsapp-key-to-quickly-rallying-protesters-in-hong-kong-but-groups-struggle-to-stay-on-message-1412878808; https://www.internetsociety.org/resources/doc/2021/factsheet-how-encryption-can-protect-advocacy-groups-and-social-change-movements/.

16 https://techagainstterrorism.org/news/2023/01/11/terrorist-use-of-end-to-end-encryption-insights-from-a-year-of-multi-stakeholder-discussion

17 https://www.washingtonpost.com/politics/2024/02/28/supreme-court-revives-debate-over-social-media-public-square/

18 https://www.economist.com/interactive/the-world-ahead/2023/11/13/2024-is-the-biggest-election-year-in-history. Among the elections occurring in 2024 are the world's largest democracy – India – and the world's richest – the United States. Several elections are taking place in countries that are considered backsliding democracies – such as Hungary, Mexico, Turkey, Tunisia, and South Africa.

19 https://engage.sinch.com/blog/most-popular-messaging-apps-in-the-world/#Global

20 Riedl, M. J., Ozawa, J. V. S., Woolley, S., & Garimella, K. (October 2022). Talking Politics on WhatsApp: A survey of Cuban, Indian, and Mexican American diaspora communities in the United States. *Center for Media Engagement*. https://mediaengagement.org/research/whatsapp-politics-cuban-indian-mexican-american-communities-in-the-united-states/

21 https://engage.sinch.com/blog/whatsapp-in-the-us-potential/. See also, https://www.androidheadlines.com/2024/05/whatsapp-double-digit-growth-in-the-us.html.

22 Hendrix, J., Quentin, C., Sinders, C., Wagner, L.W., Bernard, T., & Mehta, A. (June 2023). What is Secure? Analysis of Popular Messaging Apps. *Tech Policy Press*. https://cdn.sanity.io/files/3tzzh18d/production/249bacf0c26005325181333271be32e92024e0e5.pdf

23 See also Puyosa, I. (August 2023). Protecting point-to-point messaging apps: Understanding Telegram, WeChat, and WhatsApp in the United States. Atlantic Council. https://www.atlanticcouncil.org/in-depth-research-reports/report/point-to-point-messaging-apps/

24 Notably, LINE is not explicitly referenced in this report because of its heavy user-base concentration in East Asia and limited relevance in parts of the world where field research was conducted. iMessage is also not included in this report because it offers limited functionalities and features, beyond plain texting, that can be exploited by propagandists.

25 https://about.fb.com/news/2023/12/default-end-to-end-encryption-on-messenger

26 https://blog.cryptographyengineering.com/2024/08/25/telegram-is-not-really-an-encrypted-messaging-app/

27 See appendix II (survey) question 5. Participants rated Telegram as "very secure" more often than Viber, even though Viber offers end-to-end encryption across more features on its service, and only slightly (6 percentage points) less often than Signal despite the latter app being significantly more privacy protective.

28 The type of encryption Telegram applies to the majority of its features is client-server encryption, which is a partial form of encryption that encrypts content only between each end of the communication and the company's server, rather than end-to-end encryption which protects the contents from end to end. Nevertheless, Telegram promotes its platform by stating, "We made it our mission to provide the best security combined with ease of use. Everything on Telegram, including chats, groups, media, etc. is encrypted using a combination of 256-bit symmetric AES encryption, 2048-bit RSA encryption, and Diffie–Hellman secure key exchange." https://play.google.com/store/apps/details?id=org.telegram.messenger. Such a statement is likely to confuse or mislead users who are not aware of the differences between different types of encryption. On why Telegram users should reexamine their reliance on Telegram's claims of security, see Hendrix, J., Quentin, C., Sinders, C., Wagner, L.W., Bernard, T., & Mehta, A. (June 2023). What is Secure? Analysis of Popular Messaging Apps. *Tech Policy Press*. https://cdn.sanity.io/files/3tzzh18d/production/249bacf0c-26005325181333271be32e92024e0e5.pdf

29 https://www.theguardian.com/technology/article/2024/may/18/npr-elon-musk-signal

30 With the exception of metadata and the recently launched "channels."

31 With features ranging from supergroups, called "communities," to a recently rolled out generative AI-powered chatbot that can answer all manner of queries and several affordances for businesses, WhatsApp today is a messaging-social media hybrid. See Evangelista, R. & Bruno, F. (2019). WhatsApp and political instability in Brazil: targeted messages and political radicalisation. *Internet Policy Review*, 8(4). https://doi.org/10.14763/2019.4.1434; Hendrix, J., Quentin, C., Sinders, C., Wagner, L.W., Bernard, T., & Mehta, A. (June 2023). What is Secure? Analysis of Popular Messaging Apps. *Tech Policy Press*. https://cdn.sanity.io/files/3tzzh18d/production/249bacf0c26005325181333271be32e92024e0e5.pdf; Nobre, G. P., Ferreira, C. H. G., & Almeida, J. M. (2022). A hierarchical network-oriented analysis of user participation in misinformation spread on WhatsApp. *Information Processing & Management*, 59(1), 102757. https://doi.org/10.1016/j.ipm.2021.102757; Melo, P.F., Vieira, C.C., Garimella, K., Melo, P.O., & Benevenuto, F. (2019). Can WhatsApp Counter Misinformation by Limiting Message Forwarding? *ArXiv*, https://doi.org/10.48550/arXiv.1909.08740

32 See e.g., https://www.viber.com/en/features/ ("The information you share is safeguarded by our security systems so you never have to think twice about what you can or can't share when you're using Viber...Encryption keys exist on user devices and nowhere else. So, no one—not even Viber—can read your messages."); https://www.whatsapp.com/security ("Your privacy is our priority. With end-to-end encryption, you can be sure that your personal messages stay between you and who you send them to.").

33 Interview in India on 2 April 2024.

34 This feature is available on the four apps considered here, with some variation. On WhatsApp, broadcasts reach max 256 contacts at a time); on Viber; on Signal, media can be broadcast to max 5 contacts at a time; on Telegram, there is no limit on broadcasting.

35 Interview in India on 2 April 2024.

36 Interviews in Bolivia, India, Nigeria, Turkey, Tunisia, US (online or in person) between September 2023 – July 2024.

37 Ibid.

38 Interviews in The Philippines between June 2021 – June 2022 (online).

39 Interviews in The Philippines (online), September and October 2021.

40 https://restofworld.org/2024/bjp-whatsapp-modi/

41 Interview in India on 14 March 2024. See also: https://restofworld.org/2024/bjp-whatsapp-modi/

42 Interview in India on 14 March 2024.

43 Interviews in Bolivia, India, Nigeria, Turkey, Tunisia, US (online or in person) between September 2023 – July 2024. In some contexts, such as in Mexico, propagandists expressed a preference for larger groups over communities. However, in India, propagandists fancied communities for their versatility, announcement feature, and facilitation of coordination among propagandists.

44 Ibid.

45 https://economictimes.indiatimes.com/news/elections/lok-sabha/maharashtra/i-extend-unconditional-support-to-nda-for-modi-mns-chief-raj-thackeray/articleshow/109173505.cms?from=mdr

46 Interview in India on 2 April 2024.

47 Glover, K., Dila, M., Pate, N., Little, K., Trauthig, I., & Woolley, S. (June 2023). Encrypted messaging applications and political messaging: How they work and why understanding them is important for combating global disinformation. *Center for Media Engagement*. https://mediaengagement.org/research/encrypted-messaging-applications-and-political-messaging; https://faq.whatsapp.com/1053543185312573

48 https://faq.whatsapp.com/1053543185312573?helpref=faq_content

49 See appendix II (survey) question 8.

50 Hall, N., Lawson, B.T., Vaccari, C., & Chadwick, A. (June 2023). Beyond Quick Fixes: How Users Make Sense of Misinformation Warnings on Personal Messaging. *Everyday Misinformation Report*. https://www.lboro.ac.uk/research/online-civic-culture-centre/news-events/articles/o3c-4-beyond-quick-fixes/

51 Melo, P.F., Vieira, C.C., Garimella, K, Melo, P.O., & Benevenuto, F. (2019). Can WhatsApp Counter Misinformation by Limiting Message Forwarding? *ArXiv*, https://doi.org/10.48550/arXiv.1909.08740

52 Viber used to have a forwarding limit of one message at a time, but it increased this limit in 2017: "In order to make your life easier and your experience so much more convenient, you can now forward not only one message at a time but multiple messages to multiple recipients." https://www.viber.com/en/blog/2017-06-28/forward-multiple-messages-and-more-viber-desktop/

53 https://medium.com/@tienbm/discover-the-power-of-telegram-auto-forwarding-bot-unleash-your-chat-management-potential-1f97beca52d4

54 Woolley, S. & Monaco, N. (2020). Amplify the Party, Suppress the Opposition: Social Media, Bots, and Electoral Fraud. *Georgetown Law Tech Review*, 4(2), 447-461. https://georgetownlawtechreview.org/amplify-the-party-suppress-the-opposition-social-media-bots-and-electoral-fraud/GLTR-07-2020/

55 Interviews in the US (online and in-person) between January – May 2024.

56 Interviews in the Hungary and the US between November 2023 and April 2024 (online).

57 Interviews in Bolivia, India, Nigeria, Turkey, Tunisia, US (online or in person) between September 2023 – July 2024.

58 WhatsApp announced in April 2024 a cross-posting feature between WhatsApp statuses and Instagram stories. Users will have to opt into this feature. https://timesofindia.indiatimes.com/technology/social/users-will-soon-be-able-to-share-whatsapp-status-updates-to-instagram-all-details/articleshow/109247766.cms. WhatsApp users are already able to cross post WhatsApp statuses to their Facebook profiles. https://www.thenews.com.pk/latest/1062744-how-to-crosspost-your-whatsapp-status-to-facebook-with-these-simple-steps

59 Hendrix, J., Quentin, C., Sinders, C., Wagner, L.W., Bernard, T., & Mehta, A. (June 2023). What is Secure? Analysis of Popular Messaging Apps. *Tech Policy Press*. https://cdn.sanity.io/files/3tzzh18d/production/249bacf0c26005325181333271be32e92024e0e5.pdf

60 Interview in India on 29 February 2024.

61 Interview in India on 28 February 2024.

62 Interviews in Bolivia, India, Nigeria (online or in person) between September 2023 – July 2024.

63 Interviews in India, Nigeria, Turkey, US (online or in person) between September 2023 – July 2024.

64 Interviews in Mexico in January 2024 (online).

65 The interviewees did not share how long the channel was running. They also did not know why the channel was disbanded but assumed someone reported them. According to them, they never received official communications from WhatsApp explaining the removal.

66 https://www.whatsapp.com/business/api. The API is distinct from the WhatsApp Business app, which is targeted at small businesses and has the same features as the regular WhatsApp platform. The API offers far more opportunities for interacting with customers at scale. https://business.whatsapp.com/whatsapp-vs-whatsapp-business

67 https://www.thehindu.com/elections/lok-sabha/bulk-whatsapp-messages-with-pm-letter-mcc-violation-says-congress/article67965291.ece?utm_source=substack&utm_medium=email; https://www.livemint.com/politics/news/viksit-bharat-sam-park-row-pm-modi-bjp-whatsapp-message-controversy-congress-calls-it-political-propaganda-kerala-meta-11710656704618.html. See also, https://restofworld.org/2024/bjp-whatsapp-modi/

68 https://developers.facebook.com/docs/whatsapp/messaging-limits/

69 Interviews in India, Nigeria, Tunisia (online) between September 2023 – July 2024.

70 Interviews in Nigeria in January and February 2024 (online).

71 https://www.forbusiness.viber.com/en/messaging-solutions/

72 Interviews in Ukraine between October 2021 and July 2022 (online).

73 Interviews in The Philippines between June 2021 – June 2022 (online).

74 To the question, "Did the political content come directly from a person/account you know, a person/account you don't know, or both?" these respondents answered either "person I don't know" or "both."

75 https://www.statista.com/forecasts/1308726/whatsapp-business-global-revenues-by-region#:~:text=In%202023%2C%20the%20estimated%20revenues,popular%20instant%20messaging%20communication%20app

76 https://www.usesignhouse.com/blog/viber-stats

77 https://www.ft.com/content/c70ef7d6-230a-4404-b854-2e75fe0f2e0a

78 https://telegram.org/blog/telegram-business ("Starting today, anyone can turn their Telegram account into a business account."). For the full benefits and features available to premium subscribers, see https://telegram.org/faq_premium#features

79 Interviews in India between February – June 2024 (in person and online).

80 https://telegram.org/verify. The process involves submitting evidence of "notability," which essentially just requires a showing of social media or Internet presence – both of which can be manufactured. On premium badges, see https://t.me/premium/13

81 https://telegram.org/blog/telegram-business. For example, business users can connect Telegram bots that will process and respond to messages on their behalf. Businesses can integrate existing tools and workflows, or add AI assistants to manage their chats. Businesses can configure which chats the bots have access to, for example by excluding all chats with your contacts, or selecting only new chats. Business features are free to Telegram Premium subscribers.

82 https://ads.telegram.org

83 https://ads.telegram.org/guidelines

84 Interviews in India, Nigeria, US (online and in person) between September 2023 – July 2024.

85 Trauthig, I. K., Martin, Z. C., & Woolley, S. C. (2024). Messaging Apps: A Rising Tool for Informational Autocrats. *Political Research Quarterly,* 77(1), 17-29. https://doi.org/10.1177/10659129231190932

86 Meleshevich, K. & Schafer, B. (2018). Online Information Laundering: The Role of Social Media. *Alliance for Securing Democracy*. https://securingdemocracy.gmfus.org/online-information-laundering-the-role-of-social-media/

87 Interview in Bolivia on 16 February 2024 (online).

88 https://medium.com/@investigator515/unmasking-the-illusion-exposing-the-deceptive-web-of-sock-puppet-social-media-accounts-c5e1bcedb389; https://www.gendigital.com/blog/archive/identifying-sockpuppet-accounts-social-media

89 https://theconversation.com/social-media-apps-have-billions-of-active-users-but-what-does-that-really-mean-226021#:~:text=-Factors%20such%20as%20bot%20accounts,to%20truly%20understand%20online%20audiences.

90 Interviews in Brazil, Indonesia, Ethiopia and Eritrea, Mexico, Myanmar between October 2021 – May 2022.

91 https://fragment.com

92 Telegram X is a sleeker, faster, and more battery-efficient version of Telegram launched by Telegram in January 2018 that offers smoother animations, different swiping actions, and a cleaner, optimized layout. In March 2018, an update allowed Telegram X users to sign into as many accounts as desired and seamlessly switch between accounts, with different chat themes marking different accounts. https://telegram.org/blog/telegram-x?setln=nl; https://telegram.org/blog/telegram-x

93 Interviews in India, Turkey, Tunisia (online and in person) between September 2023 – July 2024.

94 Interviews in India (online and in person) between February – June 2024. Details on this tactic are omitted to avoid inadvertently aiding bad actors. Researchers, policymakers, and companies can reach out to authors of this report for additional explanation.

95 Interviews in India (online) on 20 Janaury 2021.

96 https://www.washingtonpost.com/world/2023/09/26/hindu-nationalist-social-media-hate-campaign/; https://firstmonday.org/ojs/index.php/fm/article/view/13172

97 https://restofworld.org/2024/bjp-whatsapp-modi/

98 Interviews in India (online and in person) between February – June 2024.

99 Monaco, N. & Nyst, C. State-Sponsored Trolling: How Governments Are Deploying Disinformation as Part of Broader Digital Harassment Campaigns. *Institute for the Future*. https://legacy.iftf.org/statesponsoredtrolling/

100 Glover, K., Dila, M., Pate, N., Little, K., Trauthig, I., & Woolley, S. (June 2023). Encrypted messaging applications and political messaging: How they work and why understanding them is important for combating global disinformation. *Center for Media Engagement*. https://mediaengagement.org/research/encrypted-messaging-applications-and-political-messaging

101 https://core.telegram.org/bots

102 https://help.viber.com/hc/en-us/articles/8746671603485-Chatbot-Commercial-Model-Legacy

103 https://help.viber.com/hc/en-us/articles/15247629658525-Bot-commercial-model

104 https://faq.whatsapp.com/666225138813752/?cms_platform=web

105 Interviews in India, Nigeria, Tunisia (online and in person) between September 2023 – July 2024.

106 https://support.signal.org/hc/en-us/articles/360007319251-Dual-Sim

107 Interview with Riana Pfefferkorn.

108 Interviews in India (online and in person) between February – June 2024.

109 Interview in Ukraine (online) on 19 August 2021.

110 Interviews in India (online and in person) between February – June 2024.

111 https://gifct.org/

112 Kazemi, A., Garimella, K., Shahi, G.K., Gaffney, D., Hale, S.A. (2022). Research note: Tiplines to uncover misinformation on encrypted platforms: A case study of the 2019 Indian general election on WhatsApp. *Harvard Kennedy School Misinformation Review*. https://misinforeview.hks.harvard.edu/article/research-note-tiplines-to-uncover-misinformation-on-encrypted-platforms-a-case-study-of-the-2019-indian-general-election-on-whatsapp/

113 https://www.theinformation.com/articles/meta-cuts-fact-checking-efforts-on-whatsapp-as-elections-loom?rc=tltwje

114 https://faq.whatsapp.com/5059120540855664. See also, Michael Rein, https://www.techpolicy.press/give-group-admins-tools-to-fight-disinformation-in-immigrant-diaspora-whatsapp-groups/ ("On WhatsApp, certified publications can be set up as bots for users to chat with and share links, videos, and memes for review.).

115 For example, WhatsApp should consider allowing users to click on a message and then be offered the option to send the message to a tipline of their choice. Interviews with Ashkan Kazemi and Kiran Garimella.

116 To contact a tipline, users need to add the tipline's phone number to their directory. Then, they must find the contact and send a message with the information. This process takes significant time and effort. Furthermore, most users are not aware that tiplines even exist. Interviews with Kiran Garimella and Ashkan Kazemi.

117 https://newsinitiative.withgoogle.com/resources/trainings/fundamentals/reverse-image-search-verifying-photos/

118 https://faq.whatsapp.com/1183938229003432/?helpref=hc_fnav

119 "This feature is currently available for users in Argentina, Brazil, Chile, Colombia, Dominican Republic, France, Germany, Ireland, Italy, Mexico, Nigeria, Peru, Spain, United Kingdom, United States, and Venezuela." https://faq.whatsapp.com/1183938229003432/?helpref=hc_fnav

120 Reis, J.C., Melo, P.F., Garimella, K., & Benevenuto, F. (2020). Detecting Misinformation on WhatsApp without Breaking Encryption. *ArXiv*, abs/2006.02471. See also, Scheffler, S., & Mayer, J. (2023). SOK: Content moderation for end-to-end encryption. *Proceedings on Privacy Enhancing Technologies*, 2023(2), 403–429. https://doi.org/10.56553/popets-2023-0060

121 Interviews with Sarah Scheffler and Kiran Garimella.

122 See, e.g., Instagram's nudity protection. https://about.instagram.com/blog/announcements/new-tools-to-help-protect-against-sextortion-and-intimate-image-abuse#:~:text=When%20nudity%20protection%20is%20turned,they've%20changed%20their%20mind.

123 Lanius, C., Weber, R. & MacKenzie, W.I. Use of bot and content flags to limit the spread of misinformation among social networks: a behavior and attitude survey. *Soc. Netw. Anal. Min.* 11, 32 (2021). https://doi.org/10.1007/s13278-021-00739-x; Gaozhao, D. (2021). Flagging fake news on social media: An experimental study of media consumers' identification of fake news. *Government Information Quarterly*, 38(3), 101591. https://doi.org/10.1016/j.giq.2021.101591

124 https://www.chicagobooth.edu/review/can-ai-stop-fake-news

125 Interview with Lucy Qin. Potential downsides include the risks of autocratic hijacking, poisoning attacks, and user overreliance on the matching.

126 Interview with Riana Pfefferkorn.

127 Pfefferkorn, R. (2022). Content-Oblivious Trust and Safety Techniques: Results from a Survey of Online Service Providers. *Journal of Online Trust and Safety*, 1(2). https://doi.org/10.54501/jots.v1i2.14. Metadata includes "phone numbers of senders and recipients; timestamps of the messages; IP addresses or other connection information; sender and recipient's mobile carriers; protocol headers, such as User-Agent strings which may contain device manufacturers and models; whether the message has an attachment; the URL on the content server where the attachment is stored; and the approximated size of messages" (Hendrix et al, 2023).

128 https://techcrunch.com/2017/02/02/how-whatsapp-is-fighting-spam-after-its-encryption-rollout/

129 In the case of Viber, the company has a spam-detection system that relies on user feedback, rather than on metadata. Viber users can turn on automatic spam-check to allow Viber to scan messages from "people that you never communicated with on Rakuten Viber." The spam check scans messages that contain email addresses, links, and phone numbers and detect if they contain malicious content from suspected scammers. https://help.viber.com/hc/en-us/articles/8923936997661-Spam-and-chain-messages

130 https://www.justsecurity.org/10311/michael-hayden-kill-people-based-metadata/

131 https://www.whatsapp.com/privacy

132 https://www.theverge.com/23409716/signal-encryption-messaging-sms-meredith-whittaker-imessage-whatsapp-china

133 https://www.nytimes.com/2021/01/13/technology/whatsapp-data.html

134 See appendix II (survey) question 17.

135 Meisner, C. (2023). The weaponization of platform governance: Mass reporting and algorithmic punishments in the creator economy. *Policy & Internet*, 15(4), 466–477. https://doi.org/10.1002/poi3.359

136 https://faq.whatsapp.com/414631957536067/?locale=en_US ("WhatsApp receives the last five messages sent to you by the reported user or group, and they won't be notified. WhatsApp also receives the reported group or user ID, information on when the message was sent, and the type of message sent (image, video, text, etc.).").

137 For more, see Center for Democracy & Technology (2021). Outside Looking In: Approaches to Content Moderation in End-to-End Encrypted Systems. *Center for Democracy & Technology*. https://cdt.org/insights/report-outside-looking-in-approaches-to-contentmoderation-in-end-to-end-encrypted-systems/; Pfefferkorn, R. (2022). Content-Oblivious Trust and Safety Techniques: Results from a Survey of Online Service Providers. *Journal of Online Trust and Safety*, 1(2). https://doi.org/10.54501/jots.v1i2.14; Lo, K. & Vilk, V. (2023). Shouting into the Void: Why Reporting Abuse to Social Media Platforms Is So Hard and How to Fix It. *PEN America*. https://pen.org/report/shouting-into-the-void/

138 Interview with Riana Pfefferkorn.

139 https://www.techpolicy.press/a-conversation-with-meredith-whittaker-president-of-signal/

140 For similar recommendations, see https://www.medianama.com/2024/04/223-mozilla-whatsapp-changes-election-disinformation/

141 See, https://www.techpolicy.press/five-eyes-campaign-against-encryption-threatens-democracy/. See also, Trauthig, I.K. (2023). Chat Apps, Mass Mobilization, and Authoritarian Control: Assessing Evidence from Egypt, Iran, and Morocco. In F. Fukuyama & M. Schaake (Eds.), *Digital technologies in emerging countries* (pp. 9-28). Stanford Cyber Policy Center. https://fsi9-prod.s3.us-west-1.amazonaws.com/s3fs-public/2023-05/digital_technologies_in_emerging_countries_digital.pdf; On the risks of such legislation in the U.S., see https://www.techpolicy.press/threatening-encryption-senate-democrats-aid-gop-war-on-abortion/

142 https://cdt.org/insights/traceability-under-brazils-proposed-fake-news-law-would-undermine-users-privacy-and-freedom-of-expression/

143 Kumar, S. (2023). Traceability and end-to-end encryption: An analysis of India's intermediary rules mandating traceability. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4344579

144 https://www.techpolicy.press/indonesias-intermediary-regulation-imperils-internet-freedom/

145 https://www.bbc.com/news/technology-67221691

146 Malhotra, P. (2020). A Relationship-Centered and Culturally Informed Approach to Studying Misinformation on COVID-19 *Social Media + Society*, 6(3). https://doi.org/10.1177/2056305120948224

147 Kumleben, M., Woolley, S., & Joseff, K. (2022). Electoral Confusion: Contending with Structural Disinformation in Communities of Color. https://protectdemocracy.org/wp-content/uploads/2022/06/electoral-confusion-contending-with-structural-disinformation-in-communities-of-color.pdf

148 https://sites.brown.edu/informationfutures/2024/07/06/to-fight-bad-information-a-project-taps-trusted-messengers-in-immigrant-communities/

149 https://www.techpolicy.press/whatsapp-misinformation-and-latino-political-discourse-in-the-u-s/

150 On the benefits of prebunking, see https://www.cbsnews.com/news/how-prebunking-misinformation-works-60-minutes/

151 For a positive example of such an effort, see the partnership between WhatsApp and NASSCOM foundation. https://www.medianama.com/2019/03/223-whatsapp-nasscom-fake-news/, and the partnership with Soy Digital in Mexico, https://wethink-digital.fb.com/mx/es-mx/

152 https://www.techpolicy.press/give-group-admins-tools-to-fight-disinformation-in-immigrant-diaspora-whatsapp-groups/

153 See, e.g., Badrinathan, S. (2021). Educative Interventions to Combat Misinformation: Evidence from a Field Experiment in India. American Political Science Review, 115(4), 1325–1341. doi:10.1017/S0003055421000459

154 Garimella, K. & Chauchard S. (2024). WhatsApp Explorer: A Data Donation Tool To Facilitate Research on WhatsApp, Arxiv, https://arxiv.org/abs/2404.01328

155 Noy, C. (2008). Sampling Knowledge: The Hermeneutics of Snowball Sampling in Qualitative Research. *International Journal of Social Research Methodology*, 11(4), 327–344. https://doi.org/10.1080/13645570701401305; Howard, P. N. (2020). *Lie Machines: How to Save Democracy from Troll Armies, Deceitful Robots, Junk News Operations, and Political Operatives*. Yale University Press. https://doi.org/10.2307/j.ctv10sm8wg

156 Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology,* 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa; Flick, U., Kardorff, E. von, & Steinke, I. (2004). *A Companion to Qualitative Research*. Sage Publications. Starks, H., & Brown Trinidad, S. (2007). Choose your method: A comparison of phenomenology, discourse analysis, and grounded theory. *Qualitative Health Research*, 17(10), 1372–1380. https://doi.org/10.1177/1049732307307031

157 Dynata is the world's largest first-party data collection company, with a global reach of nearly 70 million consumers and business professionals. See https://www.dynata.com.

NYU Stern Center for Business and Human Rights
Leonard N. Stern School of Business
44 West 4th Street, Suite 800
New York, NY 10012
+1 212-998-0261
bhr@stern.nyu.edu
bhr.stern.nyu.edu

© 2024 NYU Stern Center for Business and Human Rights
All rights reserved. This work is licensed under the
Creative Commons Attribution-NonCommercial 4.0
International License. To view a copy of the license,
visit http://creativecommons.org/licenses/by-nc/4.0/.

Center for Business
and Human Rights

The University of Texas at Austin
Center for Media Engagement
Moody College of Communication