Foreign Policy at BROOKINGS

# COUNTERING DISINFORMATION AND PROTECTING DEMOCRATIC COMMUNICATION ON ENCRYPTED MESSAGING APPLICATIONS

JACOB GURSKY AND SAMUEL WOOLLEY

JUNE 2021

## EXECUTIVE SUMMARY

Encrypted messaging applications (EMAs) that rely on end-to-end encryption (E2EE), like Signal, Telegram, and WhatsApp, offer a level of intimacy and security that have made them remarkably popular among activists and others who want to communicate without fear of government surveillance. These qualities also make them a useful vector for disinformation: they offer a means of spreading untraceable claims to users via trusted contacts in a secure environment. This policy brief argues that successfully countering disinformation on EMAs does not require undermining this stronger form of encryption.

Although EMAs typically end-to-end encrypt the content of private messages, they often do not encrypt the metadata of those messages. Interventions based on that metadata show particular promise. Metadata-based forwarding limits on WhatsApp, for instance, appear to have slowed the proliferation of disinformation in India and elsewhere. Third-party evaluations of such approaches are needed to develop and guide best practices for use on other platforms, particularly given criticism of, and broader worry surrounding, WhatsApp's use of said metadata.

Many EMAs offer channels for mass public broadcasts in addition to private messaging. By building automated tools to monitor the flow of disinformation between mainstream platforms and public channels on EMAs, counter-disinformation operations can craft targeted cross-platform interventions. This is in line with the global push for increased accountability, transparency, and accessibility of tech platforms to academics and journalists.

Disinformation campaigns on EMAs are successful primarily because of the intimacy and trust they afford. Regulatory responses to disinformation EMAs should therefore target how that trust is leveraged, rather than EMAs' use of E2EE. For example, stricter advertising disclosure laws would prevent "influence farms" coordinating on EMAs from spreading untraceable political messaging.

# INTRODUCTION

Encrypted messaging applications (EMAs) that rely on end-to-end encryption (E2EE) have gained widespread adoption in recent years, with two of the largest, WhatsApp and Telegram, claiming to have surpassed two billion and 500 million users respectively.[1] WhatsApp, the most popular EMA, is a part of the Facebook ecosystem. Signal and Telegram, which exist as their own companies, were among the most downloaded apps in early 2020.[2] The security of communication which these apps offer is only one of the reasons they are massively popular — they also provide access to free international messaging, they offer easy construction of group chats, and they boast easy messaging across various phone brands and telecoms carriers. On the political front, EMAs like Signal, Telegram, and WhatsApp have not just been a boon for the human rights activists and dissident groups that depend on them for securely communicating outside of government surveillance. Malicious actors also use them in attempts to manipulate public opinion and demobilize opposition.

In the United States, ahead of the 2020 presidential election, the Latino community in Florida and the South Asian community in North Carolina — both key swing states — were barraged with conspiracy-laden political messaging pushed through WhatsApp.[3] Meanwhile, rampant COVID-19 disinformation targeted the Korean-American community in Los Angeles through the ubiquitous Korean app KakaoTalk.[4] As we and co-authors have documented in a University of Texas at Austin Center for Media Engagement report, political disinformation on WhatsApp has likewise become commonplace in Mexico and India and is often sanctioned by or traced back to government or party officials.[5] EMAs are now regularly used to spread disinformation around the world, often in ways that hyper-target minority communities, exacerbate existing political and social tensions, and even inflame violence.

**On the one hand, EMAs are widely used by pro-democracy activists around the world and empower citizens to communicate freely without fear of government surveillance. On the other, they present new opportunities for malicious actors to engage in online harms such as disinformation campaigns.**

For policymakers concerned with disinformation, EMAs pose a unique challenge. On the one hand, EMAs are widely used by pro-democracy activists around the world and empower citizens to communicate freely without fear of government surveillance. On the other, they present new opportunities for malicious actors to engage in online harms such as disinformation campaigns. Worse, because the users of applications like Telegram tend to know each other, or be friends of friends, they are more likely to believe the disinformation they are exposed to on those apps. EMAs thus represent both a more secure vector for disinformation and a more effective one too. The disinformation is typically presented in frequently-forwarded and mass-forwarded messages sent directly from a more trusted set of contacts than on other social media platforms. The social capital in tightly knit communities on encrypted spaces creates a false sense of security that information can be trusted, while the functionality of the platforms makes source verification and attribution difficult or impossible.

Given their reliance on end-to-end encryption, the most intuitive way to counter disinformation on EMAs is to mandate a backdoor into EMAs for law enforcement. Researchers have yet to show that this is possible without increasing vulnerability to malicious actors, and the Obama administration ultimately recommended against this approach. This is part of an ongoing debate between privacy advocates and the government that has been going on since the 1990s, known as the "Crypto Wars."[6] The Trump administration, however, re-initiated the backdoor approach.[7] The bipartisan EARN IT Act in

the United States, for instance, targets online child sexual exploitation but has wider ramifications for encryption.[8] While centered on preventing child exploitation, the bill encourages the creation of backdoors by holding EMAs accountable for illegal content spread on their platforms, even if they have no way to access it because of E2EE. Even after attempts to reform the bill to "protect" encryption, it stands to encourage backdoors and discourage platforms from adopting E2EE in the first place to avoid precarious legal positions.[9] The Council of the European Union's resolution on encryption adopted in late 2020[10] has been interpreted as a similar stance by some.[11] Yet such a view is misguided. Not only does undermining E2EE pose a danger to human rights activists and dissidents — during Beijing's recent crackdown in Hong Kong, for instance, downloads of Signal spiked there[12] — but there are also more effective counter-disinformation interventions and policy responses available. Increased journalistic and academic access to the prominent public channels hosted on EMAs would be an invaluable tool in tracing disinformation and would require no damage to encryption regimes. Additionally, applying existing disclosure frameworks for political advertisements to EMAs would counter disinformation and create accountability by recognizing that disinformation exists in a large, cross-platform reality, of which EMAs are only one part.

As Facebook and other platforms make end-to-end encryption standard, these and other responses will only become more important. What is needed are policy frameworks that avoid both over-regulation that infringes on people's right to privacy and free speech, and also under-regulation that results in the spread of more — and more damaging — disinformation and harassment.

## END-TO-END ENCRYPTION, METADATA, AND THE RISE OF EMAS

Developing effective responses to disinformation campaigns on EMAs first requires an understanding of different encryption regimes. Debates over E2EE hinge on *who* has the ability to read a message and *when*. If a user is using a messaging service with no encryption at all, not only can they and their recipient read their message, but so too can anyone with access to the messaging service's servers, as well as anyone with access to the servers the message is routed through as it transits the internet. If a user is using a messaging service like Discord, which relies on transport layer security (TLS), the message is encrypted first as it travels from the user to the messaging service's server, and then again as it travels from the service's server to the recipient. As a result, the message is secure from actors (governmental or otherwise) intercepting messaging in transit between the sender and service or between the service and receiver. Yet the messaging service itself can still read it, and law enforcement and other government agencies may be able to compel the service to provide the message on request. With messaging services that use E2EE, however, neither the service nor law enforcement are able to view the message, since it remains encrypted even on the service's servers. Only the reader and recipient are able to view and read the message.

**While E2EE platforms are known for their ability to support private messaging, some, particularly WhatsApp, are known to collect metadata, or data related to — but not containing — the messages sent over the platform.**

Although most EMAs encrypt the content of a given message, they typically do not encrypt the metadata related to that message. To understand metadata

and encryption, think of a physically posted letter. The contents inside the envelope are what is encrypted, while the information on the outside of the envelope is the metadata: names, locations, data stamps, etc. The collection of metadata in apps is similar. WhatsApp collects metadata not only on the senders and recipients of a given message, but also on the frequency of messaging. While E2EE platforms are known for their ability to support private messaging, some, particularly WhatsApp, are known to collect metadata, or data related to — but not containing — the messages sent over the platform. For example, WhatsApp cannot know the call's content but could collect the identity of both parties and the conversation's initialized time and spent time. In fact, an app can collect much more user metadata including phone model, operating system, contacts, location, and so on. After professional analysis of this metadata, one can acquire a great deal of information. This can be used both to detect and counter disinformation campaigns. For instance, based on the metadata it stores, WhatsApp was able to impose forwarding restrictions in April 2020 that it claimed curbed viral messaging by 70%.[13]

The proliferation of smartphones and the ease of app store installs meant that the complexity of most E2EE systems could be abstracted away for the average user. The result has been an explosion of EMAs that rely on E2EE, such as Telegram, Signal, and WhatsApp. The reach of these platforms has only increased over time, with spikes in adoption often tied to political events. WhatsApp, founded in 2009, surpassed two billion users last February.[14] Telegram and Signal both saw a huge uptick in use after the January 6, 2021 assault on the U.S. Capitol, with Telegram gaining 25 million users in a space of 72 hours to reach over 500 million active users.[15]

As noted above, the widespread adoption of EMAs means that billions of people can now communicate securely. Yet that number is poised to grow. Firms like Facebook are seeing a large-scale shift among a userbase that mainly interacts with their ecosystem of products through E2EE. WhatsApp and Messenger exercise E2EE, and Facebook's other platforms are moving in that direction. Founder and CEO Mark Zuckerberg has, in fact, described the company's future as "privacy-focused."[16] Facebook is attempting to integrate its proprietary digital ecosystems, bringing together WhatsApp, Instagram, Messenger, and Facebook. However, as these mergers and acquisitions occur, they create problems. If E2EE is offered in one part of a company's ecosystem, it becomes expected across all, especially as messaging and sharing information within the ecosystem becomes more fluid. Google has rolled out E2EE on its Android messaging services[17] as it has become the expectation of users.

Significantly, the shift to E2EE poses a major challenge in light of disinformation operations. Because E2EE platforms lack visibility into the content of messages, it may at first glance make it seem like the problem has gone away. Yet that clearly isn't true. As we have documented, EMAs are used worldwide to, among other things, spread and organize influence across platforms, spread disinformation that has led to mob violence and deaths, provide a home for violent fringe organizations seeking to manipulate mainstream media, and give platforms to organizations spreading dangerous lies about public health, especially in relation to the COVID-19 pandemic.[18]

## COUNTERING DISINFORMATION ON EMAS

The goal of this policy brief is not just to describe the growth of misinformation and disinformation on EMAs, but to identify ways both the tech sector and policy community can effectively respond in order curb the spread of harmful behavior without dismantling or undermining E2EE, which has clear utility for everyday communication and for democratic organizing and engagement.

We recommend shifting the focus of both industry and policy responses from securing access to encrypted content towards identifying and disincentivizing cross-platform actors, who utilize EMAs as one step in a larger flow of disinformation into mainstream media. Based on our work at the Center for Media Engagement, and supported by the work of others, such as Camille François, chief innovation officer at cybersecurity company Graphika and an affiliate of Harvard University's Berkman Klein Center for Internet and Society, regulatory efforts that focus disproportionately on content on any single platform — ignoring the motivations of actors and their adaptations over time — not only bolster attempts to undermine encryption, but are also ineffective in countering disinformation and its unique challenges.[19]

It is important to recognize that coordinated deceptive behavior on EMAs is not limited to a single platform, and that the technical capabilities of E2EE are not the only incentive for actors spreading false content to utilize encrypted spaces. As discussed above, the higher levels of trust and social capital generated by EMA groups are attractive to disinformation campaigns seeking to abuse that trust, an issue which persists regardless of the specific encryption standard used. Furthermore, there are three different categories of EMA users to consider: political disinformation campaigns, the criminal or terrorist groups most often cited by law enforcement as a reason to break E2EE, and individual users concerned with their privacy. Stopping disinformation on EMAs requires a different framework to regulating encrypted spaces than the one used by law enforcement agencies in the ongoing "Crypto Wars" to counter abusive content, and it is a framework that can be built without dismantling E2EE's legitimate uses.

### *Tech interventions*

As tech giants move to consolidate multi-platform ecosystems, and as users come to expect E2EE as industry standard for messaging applications, pressure on tech companies to counter disinformation in encrypted spaces has increased. Content-agnostic approaches that utilize platform access to metadata about messages have shown some promise. WhatsApp's aforementioned implementation of forwarding limitations to curb virality is the largest public attempt to date to curb disinformation on an EMA, but due to Facebook's lack of transparency about its internal operations, it is hard to gauge the effort's true efficacy and make an assessment about whether it is a viable strategy for other platforms. WhatsApp has taken an approach to curbing disinformation that draws lines based on user behavior and metadata, and reportedly has achieved a level of success without compromising encryption.[20] However, while content-agnostic forwarding limitations may seem like a middle ground solution, it is more likely that companies will use it to dodge responsibility while continuing to reap massive profits from viral disinformation that they can argue is encrypted from their eyes, too. This becomes particularly potent as platforms combine E2EE with their attempts to consolidate their ecosystems and monopolistic positions.[21] While users deserve privacy from the platforms themselves, platforms need to make the effort to adopt more diverse, E2EE-compatible approaches, such as Camille François's "Actors, Behaviors, Content" model,[22] rather than simply curbing mass virality and calling the job done.

**It is a misconception that EMAs are entirely inaccessible to outside observers. Part of what makes them such fertile grounds for disinformation is that they allow for mass messaging groups that are highly public and easy to join.**

One promising approach is that researchers can leverage content that is publicly available on EMAs to track disinformation. Because malicious groups' primary intention is often to amplify content in public channels with a larger reach than EMAs,[23] those seeking mass reach need a channel that is accessible to those that will act as amplifiers, whether it be a public Facebook group or a WhatsApp

or Telegram broadcast channel. This leaves an opening for intervention and tracing content.[24] It is a misconception that EMAs are entirely inaccessible to outside observers. Part of what makes them such fertile grounds for disinformation is that they allow for mass messaging groups that are highly public and easy to join. Because these groups are public, it is possible to analyze disinformation in transit between EMAs and public platforms — even professional news outlets. For example, Gab is a social network popular with the far right that has shifted to E2EE,[25] where forms of harmful content are actively created and aggregated with the purpose of spreading or linking them — in a strategized and organized format — onto mainstream platforms like Facebook or YouTube.[26] First Draft, a prominent group working to protect internet users from mis- and disinformation, calls this the "Trumpet of Amplification"[27]: content is actively spread, and designed to spread, from closed to increasingly open and larger networks, ultimately ending up in the mainstream media. Additionally, Ben Nimmo's "Breakout Scale" is an example of a framework that moves beyond the technical capabilities of any individual platform and instead focuses on cross-platform spread and amplification of malicious content.[28]

To facilitate analysis of such transmission, the tech sector and research community should invest in developing tools that identify and catalogue the ways in which actors circumvent regulations and spread content through analysis of the content itself. For example, machine learning tools that can identify a viral screenshot on Twitter as having originated on an EMA based upon its background, and then cross-reference the screenshot with datasets of public groups scraped from EMAs, would be useful. From there, the administrators of the groups can be investigated, including their relationships with private chats that are completely hidden. For example, a dataset of public WhatsApp groups in Brazil called WhatsApp Monitor,[29] maintained by Professor Fabricio Benevenuto and his team at the Federal University of Minas Gerais, is a point of reference for curbing and understanding mis- and disinformation there. While such datasets do not collect private E2EE chats, they provide the opportunity to trace the amplified messaging from mainstream platforms back to EMAs, at which point targeted interventions, that do not rely on compromising encryption for legitimate users, can be used to identify actors who are leveraging smaller E2EE chats on the same platform to originate deceptive content.

The value of the public channel monitor approach also has potential for disinformation that spreads solely on EMAs, as is often the trend in India, Brazil, and Mexico — as well as those country's diaspora communities in the United States and elsewhere. Creating usable datasets of Telegram[30] and WhatsApp[31] public groups has already been proven feasible, and investing time, energy, and resources into developing such tools can help curb harmful content without damaging the security of the platforms. Even as firms like Facebook seek to make E2EE the default across their increasingly consolidated app ecosystems, the sorts of interventions described above will be useful because they focus on the behavior of the actors rather than the content. Centralized datasets of illegal and extremist content in a hashed form for cross-platform reference has already been proven feasible by the Global Internet Forum to Counter Terrorism, and is being viewed as a viable path forward for countering domestic white supremacy and terrorism in the wake of the attack on the U.S. Capitol on January 6, 2020.[32] However, these databases of hashes are only feasible for monitoring public channels and groups on EMAs; applying them to private chats and groups undermines the promises of E2EE.[33] Unlike criminals sharing plans or images that must be kept hidden from authorities, disinformation spreaders rely on the ability to share messages with a wide audience. Platforms adopting policies that encourage journalist and researcher access to content from public channels, in order to facilitate cross-platform tracing of disinformation, would strengthen counter-disinformation efforts.

## *Policy interventions*

Since EMAs rely on end-to-end encryption, any online harms they facilitate — such as disinformation — are often met with policy proposals to undermine the encryption. The EARN IT Act,[34] introduced March 5, 2020, largely focused on curbing the spread of child abuse content, is receiving criticism as an attack on encryption.[35] In addition to the EARN IT Act, the Lawful Access to Encrypted Data Act, introduced on June 23, 2020, seeks to "require service providers and device manufacturers to provide assistance to law enforcement when access to encrypted devices or data is necessary — *but only after* a court issues a warrant, based on probable cause that a crime has occurred, authorizing law enforcement to search and seize the data."[36] It has received similar criticisms.[37]

The Electronic Privacy Information Center (EPIC)'s Interim Executive Director Alan Butler said the latter bill "will make it easier for bad actors to access people's communications. You cannot build a backdoor that only law enforcement can access. That's not how encryption works."[38] EPIC earlier told the Senate Judiciary Committee "now is not the time to undermine the systems that we all rely upon to secure our data and communications."[39] There is no way to weaken or provide exceptional access to an encrypted system without making it weaker for everyone.

**Legislation designed to undermine encryption standards is not the best way to respond to disinformation and other harms posed by EMAs, because bad actors can exploit the same backdoors that legislators are trying to install for oversight and law enforcement purposes.**

Legislation designed to undermine encryption standards is not the best way to respond to disinformation and other harms posed by EMAs, because bad actors can exploit the same backdoors that legislators are trying to install for oversight and law enforcement purposes. However, leaving these spaces unregulated altogether is also problematic, since it leaves them open to the spread of coordinated, large-scale disinformation and political manipulation campaigns. The current approaches to regulation are, we feel, agnostic to the motivations and tactics of those spreading disinformation and instead are focused on the technology itself. Regulations should be more focused on actors. Such actor-oriented policy can be stringent while maintaining the integrity of the technology.

We argue for two policy responses to misinformation and disinformation on EMAs.

First, in line with above, we encourage policymakers to recognize that regulations targeting disinformation campaigns do not need to focus on dismantling E2EE. Rather, they should focus on making it easier for researchers and journalists to have access to the public data needed to approach the disinformation ecosystem for what it is — a cross-platform ecosystem in which the public channels on EMAs are important and where the motivations and specific stratagems of actors are crucial.

The following example from our ongoing research highlights how many strategies on closed messaging platforms do not rely on encryption to seed and fertilize their messages. In the wake of the 2018 mass shooting at Marjory Stoneman Douglas High School in Parkland, Florida, the fraudulent and opportunistic white supremacist group known as the Republic of Florida[40] organized a successful campaign on Discord in real time to convince mainstream and national news outlets that the shooter identified with their group, which he did not.[41] As mentioned earlier in the brief, Discord is not an E2EE platform — messages are only encrypted in transit. In fact, records of the group's chats are still available online in image collages on a well-known image-hosting website. It is our view that many of the current arguments around E2EE can be traced to the decades-long "Crypto Wars" over government access to encrypted communication,[42]

more so than to any particular value that E2EE lends to disinformation campaigns. The Republic of Florida example supports the argument that building regulations with a focus on cross-platform coordination and an understanding of actors' motivations is more effective than building around specific technological capabilities of the platforms. Simply compromising E2EE will not stop messaging-group disinformation, as may be seen from apps such as Discord which are similar to EMAs but do not use E2EE. Whether or not Discord used E2EE was not relevant to the effectiveness of the manipulation campaign, as it relied on coordinating offsite on one platform (Discord) to manipulate journalists on another (Instagram).

Second, advertising disclaimers for digital campaigns that rely on organized human operatives operating within their trusted communities (influencers) can curb the spread of mis- and disinformation at the point of creation. In our research at the Center for Media Engagement, we have identified a phenomenon known as Instagram "pods" in the United States that are leveraging spaces like EMAs to generate disingenuous influence within Facebook's app ecosystem — though not yet, to our knowledge, political influence. EMAs (notably Telegram[43]) are often used in these influence farming and "like sharing" operations. In these "pods," users on public channels in EMAs coordinate like and comment shares to grow their positions of influence on Instagram. Companies such as Wolf Global[44] also cultivate off-platform influence on EMAs through bot-moderated like and comments sharing. Users seeking to bolster their individual influence use the service for free, but the companies are able to use these communities to provide instant influence for clients who pay to have their content promoted. We have also identified a similar trend in Mexico, where an app used to host pods is used explicitly by politicians and known propagandists to generate false engagement on social media. These propagandists coordinate amongst themselves on the EMA Telegram.[45] We believe these pods are poised for similar political usages in the United States and that requiring stricter political disclosure of influencer campaigns on platforms like Instagram would curb cross-platform coordination of inauthentic engagement and behavior without undermining encryption.

Coordinated campaigns leveraging intimacy should have stricter disclosure laws in line with comments from Federal Election Commission Chair Ellen L. Weintraub stating the necessity of reforming political ads online:

> "Americans deserve transparency when it comes to internet communications, especially given the growing threat of online disinformation campaigns and false political advertising. The FEC needs to do its part to combat these threats and make it harder for foreign adversaries to interfere in our elections with their influence operations. Better rules for internet ads are a small but necessary step."[46]

## CONCLUSION

Reapplying frameworks for regulation developed to counter illegal and abusive content misses unique challenges to countering disinformation. While the debate around E2EE and law enforcement is more relevant than ever, especially with the increased reliance on digital communication during the pandemic, we recommend that those looking to build policy to curb mis- and disinformation on these platforms recognize that intimacy is often more important than security to those running large-scale political disinformation campaigns. As large broadcast platforms like Twitter and Facebook become more adept at countering automated mis- and disinformation, the creators and spreaders of that content are turning to EMAs, finding ways to leverage trust and intimacy through real people or through unmoderated systems.

A 2019 essay by former Federal Bureau of Investigation General Counsel Jim Baker well summarizes the democratic need for encryption while also accepting associated challenges:

> "In the face of congressional inaction, and in light of the magnitude of the threat, it is time for governmental authorities — including law enforcement — to embrace encryption because it is one of the few mechanisms that the United States and its allies can use to more effectively protect themselves from existential cybersecurity threats, particularly from China. This is true even though encryption will impose costs on society, especially victims of other types of crime… I am unaware of a technical solution that will effectively and simultaneously reconcile all of the societal interests at stake in the encryption debate, such as public safety, cybersecurity and privacy as well as simultaneously fostering innovation and the economic competitiveness of American companies in a global marketplace."[47]

Policy changes in two key areas can curb the spread of misinformation on EMAs by recognizing that mis- and disinformation in encrypted spaces is different from abusive and illegal content as it is defined in legislation like the EARN IT Act. The first is to reform political influence disclosure practices on platforms.[48] Misinformation does not recognize borders between platforms the way regulators do, and requiring disclosure on very public platforms will affect the private platforms where messaging is being developed and coordinated. This would, for instance, create an obstacle for political actors seeking to use pods and other forms of like and comment sharing on EMAs to misrepresent their support on mainstream platforms like Facebook and Twitter. As far as our lab at the Center for Media Engagement can tell, there is nothing stopping political actors using EMAs from manipulating Instagram, for example, in ways invisible to Instagram users.

Second, the companies behind EMAs need to facilitate the use of publicly available content from common broadcast channels and groups to create systematic ways of tracking cross-platform disinformation and provide insight into the actors who are responsible for the closed groups on the same platforms where dangerous content originates. This would allow investigations and disruption into mis- and disinformation on EMAs to trace content from mainstream platforms back to their source platform and provide law enforcement with information into actors who can be investigated without requiring a backdoor to the platform itself. Currently, there are only a few scattered projects seeking to centralize this information, and the researchers who build them — many of whom we have interviewed in the course of our research — are met with roadblocks from the platforms themselves and forced to find creative and non-exhaustive workarounds. All of these strategies will benefit from approaching regulations with a focus on diverse actors and their equally diverse behaviors, rather than a myopic focus on the dangers of E2EE spaces.

# REFERENCES


1 Pavel Durov, "400 Million Users, 20,000 Stickers, Quizzes 2.0 and €400K for Creators of Educational Tests," Telegram, April 24, 2020, https://telegram.org/blog/400-million; "About WhatsApp," WhatsApp, https://www.whatsapp.com/about/; "Moving Chat History from Other Apps," Telegram, January 28, 2021, https://telegram.org/blog/move-history; Boris Wertz (@bwertz), Twitter, January 12, 2021, https://twitter.com/bwertz/status/1349129671774859267.

2 Sarah Perez, "Following riots, alternative social apps and private messengers top the app stores," TechCrunch, January 11, 2021, https://techcrunch.com/2021/01/11/following-riots-alternative-social-apps-and-private-messengers-top-the-app-stores/.

3 Sabrina Rodriguez and Marc Caputo, "'This is f---ing crazy': Florida Latinos swamped by wild conspiracy theories," *Politico*, September 14, 2020, https://www.politico.com/news/2020/09/14/florida-latinos-disinformation-413923; Karly Domb Sadof, "Fake news spread on WhatsApp to Indian Americans plays stealth role in U.S. election," NBC News, October 27, 2020, https://www.nbcnews.com/tech/tech-news/fake-news-spread-whatsapp-indian-americans-plays-stealth-role-us-elect-rcna166.

4 Allegra Frank and Daniel Markus, "How the coronavirus rumor mill can thrive in private group chats," Vox, March 5, 2020, https://www.vox.com/2020/3/5/21165238/coronavirus-rumors-myths-facebook-whatsapp-podcast.

5 Jacob Gursky, Katlyn Glover, Katie Joseff, Martin J. Riedl, Jimena Pinzon, Romi Geller, and Samuel Woolley, "Encrypted Propaganda: Political Manipulation Via Encrypted Messaging Apps in the United States, India, and Mexico," (Austin, TX: The University of Texas at Austin Center for Media Engagement, October 26, 2020), https://mediaengagement.org/research/encrypted-propaganda/.

6 Kristin Finklea, "Encryption and the 'Going Dark' Debate," (Washington, DC: Congressional Research Service, January 25, 2017), https://crsreports.congress.gov/product/pdf/R/R44481.

7 Martin Kaste, "Warrant-Proof Encrypted Messages Targeted By Trump Administration," NPR, February 18, 2020, https://www.npr.org/2020/02/18/806887313/warrant-proof-encrypted-messages-targeted-by-trump-administration.

8 EARN IT Act of 2020, S. 3398, 116th Cong. (2020), https://www.congress.gov/bill/116th-congress/senate-bill/3398/text.

9 Riana Pfefferkorn, "House Introduces EARN IT Act Companion Bill, Somehow Manages to Make It Even Worse," Center for Internet and Society, Stanford Law School, October 5, 2020, http://cyberlaw.stanford.edu/blog/2020/10/house-introduces-earn-it-act-companion-bill-somehow-manages-make-it-even-worse.

10 "Encryption: Council adopts resolution on security through encryption and security despite encryption," Council of the European Union, December 14, 2020, https://www.consilium.europa.eu/en/press/press-releases/2020/12/14/encryption-council-adopts-resolution-on-security-through-encryption-and-security-despite-encryption/.

11 Richie Koch, "EU's resolution on encryption foreshadows likely anti-encryption push," ProtonMail, December 15, 2020, https://protonmail.com/blog/eu-attack-on-encryption/.

12 Felix Tam, "Signal Tops Hong Kong Downloads After Fears of China Law Deepen," Bloomberg, July 8, 2020, https://www.bloomberg.com/news/articles/2020-07-08/signal-messenger-rockets-up-hong-kong-download-charts.

13 Jon Porter, "WhatApp says its forwarding limits have cut the spread of viral messages by 70 percent," The Verge, April 27, 2020, https://www.theverge.com/2020/4/27/21238082/whatsapp-forward-message-limits-viral-misinformation-decline.

14 "Two Billion Users – Connecting the World Privately," WhatsApp, February 12, 2020, https://blog.whatsapp.com/two-billion-users-connecting-the-world-privately.

15 Jack Nicas, Mike Isaac, and Sheera Frenkel, "Millions Flock to Telegram and Signal as Fears Grow Over Big Tech," *The New York Times*, January 13, 2021, https://www.nytimes.com/2021/01/13/technology/telegram-signal-apps-big-tech.html; Megha Mandavia, "Telegram crosses 500 million monthly active users globally," *The Economic Times*, January 13, 2021, https://economictimes.indiatimes.com/tech/technology/telegram-crosses-500-million-monthly-active-users-globally/articleshow/80245013.cms.

16 Mark Zuckerberg, "A Privacy-Focused Vision for Social Networking," Facebook, March 6, 2019, https://www.facebook.com/notes/mark-zuckerberg/a-privacy-focused-vision-for-social-networking/10156700570096634/.

17 Drew Rowny, "Helping you connect around the world with Messages," Google, November 19, 2020, https://blog.google/products/messages/helping-you-connect-around-world-messages/.

18 Jacob Gursky, Katlyn Glover, Katie Joseff, Martin J. Riedl, Jimena Pinzon, Romi Geller, and Samuel Woolley, "Encrypted Propaganda."

19 Camille François, "Actors, Behaviors, Content: A Disinformation ABC," (Amsterdam: Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, Institute for Information Law, University of Amsterdam, September 20, 2019), https://science.house.gov/imo/media/doc/Francois%20Addendum%20to%20Testimony%20-%20ABC_Framework_2019_Sept_2019.pdf.

20 Jon Porter, "WhatsApp says its forwarding limits have cut the spread of viral messages by 70 percent."

21 Kaya Yurieff, "Facebook takes a big step in linking Instagram, Messenger and WhatsApp," CNN, September 30, 2020, https://www.cnn.com/2020/09/30/tech/instagram-messenger-messaging/index.html; Gail Kent, "Messenger Policy Workshop: Future of Private Messaging," Facebook, April 30, 2021, https://about.fb.com/news/2021/04/messenger-policy-workshop-future-of-private-messaging/.

22 Camille François, "Actors, Behaviors, Content."

23 "Trading up the chain," Media Manipulation Casebook, https://mediamanipulation.org/definitions/trading-chain.

24 Alex Hern, "Facebook merges Messenger chat service with Instagram," *The Guardian*, September 30, 2020, https://www.theguardian.com/technology/2020/sep/30/facebook-merges-messenger-chat-service-with-instagram.

25 Andrew Torba, "AG Barr is wrong on encryption. Introducing Gab Chat: An open source encrypted messaging platform," Gab, January 31, 2020, https://news.gab.com/2020/01/31/ag-barr-is-wrong-on-encryption-introducing-gab-chat-our-open-source-encrypted-messaging-platform/.

26 Samuel Woolley, Roya Pakzad, and Nicholas Monaco, "Incubating Hate: Islamophobia and Gab," (Washington, DC: The German Marshall Fund of the United States, June 21, 2019), https://www.gmfus.org/publications/incubating-hate-islamophobia-and-gab.

27 Claire Wardle, "5 Lessons for Reporting in an Age of Disinformation," First Draft, December 28, 2018, https://medium.com/1st-draft/5-lessons-for-reporting-in-an-age-of-disinformation-9d98f0441722.

28 Ben Nimmo, "The Breakout Scale: Measuring the impact of influence operations," (Washington, DC: The Brookings Institution, September 2020), https://www.brookings.edu/research/the-breakout-scale-measuring-the-impact-of-influence-operations/.

29 WhatsApp Monitor, http://www.monitor-de-whatsapp.dcc.ufmg.br/.

30 Jason Baumgartner, Savvas Zannettou, Megan Squire, and Jeremy Blackburn, "The Pushshift Telegram Dataset," in *Proceedings of the Fourteenth International AAAI Conference on Web and Social Media* (Palo Alto, CA: The AAAI Press, June 2020), 840-847, https://www.aaai.org/Library/ICWSM/icwsm20contents.php.

31 Kiran Garimella and Gareth Tyson, "WhatsApp, Doc? A First Look at WhatsApp Public Group Data," (Ithaca, NY: arXiv, Cornell University, 2018), https://arxiv.org/abs/1804.01473.

32 Bharath Ganesh, "How to Counter White Supremacist Extremists Online," *Foreign Policy*, January 28, 2021, https://foreignpolicy.com/2021/01/28/how-to-counter-white-supremacist-extremists-online/.

33 Erica Portnoy, "Why Adding Client-Side Scanning Breaks End-To-End Encryption," Electronic Frontier Foundation, November 5, 2019, https://www.eff.org/deeplinks/2019/11/why-adding-client-side-scanning-breaks-end-end-encryption.

34 EARN IT Act of 2020.

35 Joe Mullin, "The New EARN IT Bill Still Threatens Encryption and Free Speech," Electronic Frontier Foundation, July 2, 2020, https://www.eff.org/deeplinks/2020/07/new-earn-it-bill-still-threatens-encryption-and-free-speech.

36 "Graham, Cotton, Blackburn Introduce Balanced Solution to Bolster National Security, End Use of Warrant-Proof Encryption that Shields Criminal Activity," Senate Committee on the Judiciary, June 23, 2020, https://www.judiciary.senate.gov/press/rep/releases/graham-cotton-blackburn-introduce-balanced-solution-to-bolster-national-security-end-use-of-warrant-proof-encryption-that-shields-criminal-activity.

37 Richie Koch, "The Lawful Access to Encrypted Data Act wants to ban strong encryption," ProtonMail, July 22, 2020, https://protonmail.com/blog/usa-laed-act-anti-encryption/.

38 "Lawful Access to Encrypted Data Act Weakens Encryption, Undermines Public Safety," Electronic Privacy Information Center, June 24, 2020, https://epic.org/2020/06/lawful-access-to-encrypted-dat.html.

39 Marc Rotenberg, Caitriona Fitzgerald, and Alan Butler, Letter to Lindsey Graham and Dianne Feinstein, Electronic Privacy Information Center, December 9, 2019, https://epic.org/testimony/congress/EPIC-SJC-Encryption-Dec2019.pdf.

40 "Republic of Florida (ROF)," Anti-Defamation League, 2021, https://www.adl.org/resources/backgrounders/republic-of-florida-rof.

41 Thalia Beaty and Benamin T. Decker, "Source Hacking: A Disinformation Retrospective on the Parkland Shooting," First Draft, March 6, 2018, https://medium.com/1st-draft/source-hacking-a-disinformation-retrospective-on-the-parkland-shooting-585990dbb669.

42 Kristin Finklea, "Encryption and the 'Going Dark' Debate."

43 Gabriela Barkho, "Inside Instagram Pods: The Secret Trick to Increase Your Engagement," Later, February 23, 2017, https://later.com/blog/instagram-pods/.

44 Wolf Global, https://wolfglobal.org/.

45 Jacob Gursky, Katlyn Glover, Katie Joseff, Martin J. Riedl, Jimena Pinzon, Romi Geller, and Samuel Woolley, "Encrypted Propaganda."

46 Ellen L. Weintraub, "Internet Ad Disclaimers Rulemaking Proposal," (Washington, DC: Federal Election Commission, June 13, 2019), https://www.fec.gov/resources/cms-content/documents/mtgdoc_19-26-a.pdf.

47 Jim Baker, "Rethinking Encryption," Lawfare, October 22, 2019, https://www.lawfareblog.com/rethinking-encryption.

48 Anastasia Goodwin, Katie Joseff, and Samuel Woolley, "Social media influencers and the 2020 U.S. election: Paying 'regular people' for digital campaign communication," (Austin, TX: The University of Texas at Austin Center for Media Engagement, October 2020), https://mediaengagement.org/research/social-media-influencers-and-the-2020-election.

## ABOUT THE AUTHORS

**Jacob Gursky** is a research associate at the Center for Media Engagement at the University of Texas at Austin. He graduated from the University of Pennsylvania with a bachelor's degree from the undergraduate program at the Annenberg School for Communication. He is especially interested in disinformation, digital literacy, privacy activism, and the repercussions of surveillance capitalism.

**Samuel Woolley** is a leading international expert on propaganda and digital technology, with a focus upon the production of influence campaigns using infrastructural and emerging media tools: encrypted messaging applications, bots, AI, voice-emulation, and VR. He is an assistant professor of journalism and media and fellow of the R.P. Doherty, Sr. Centennial Professorship in Communication at the Moody College of Communication at the University of Texas at Austin. Woolley is a Knight faculty fellow and program director of the Propaganda Research Lab at UT's Center for Media Engagement. He is an associate fellow at GLOBSEC and a research associate at Stanford's Project for Democracy and the Internet. He has past research appointments at the University of Oxford and University of California, Berkeley and past fellowships at Google Jigsaw, the Anti-Defamation League, and the German Marshall Fund of the United States. His most recent book is *The Reality Game: How the Next Wave of Technology Will Break the Truth* (PublicAffairs/Hachette).

## ACKNOWLEDGEMENTS

Caroline Klaff and Ted Reinert edited this paper, and Rachel Slattery provided layout.