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

Center for Media Engagement
Moody College of Communication
The University of Texas at Austin
2504 B Whitis Avenue (A0730)
Austin, TX 78712-1879

To: Ms. Irene Khan, Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, United Nations
From: Center for Media Engagement Propaganda Research Lab, University of Texas at Austin
Date: 11 July 2022
Re: Report on challenges to freedom of opinion and expression in times of armed conflict and other disturbances

Encrypted messaging and chat apps (EMAs) are integral digital infrastructures for political activists, journalists, NGOs, and free expression writ large. They can also become spaces for the spread of false and misleading information, political propaganda, calls to violence, or child sexual abuse material (CSAM). At this critical juncture in time, the politicization of content and speech regulation and fervent debates around encryption foreshadow the real possibility and risk that the end is nigh for encrypted communication.

Over the past two years, the Propaganda Research Lab at the Center for Media Engagement at the University of Texas at Austin researched the spread of false and misleading information on EMAs across the globe, in countries that span geographic, political, and ideological differences: Egypt, Ethiopia, Eritrea, India, Indonesia, Libya, Mexico, Morocco, Myanmar, the Philippines, Turkey, and the United States. Based on our research, we offer insight into some of the harms that can emanate within encrypted spaces, while at the same time outlining levers of change for companies as well as platform users which might be helpful in mitigating harms while still promoting free expression through encryption. Ultimately, free expression is the cornerstone of healthy societal discourse and the basis for a political system that accommodates the needs of the many, not the elite few.

Endangering such discourse are four successful tactics that have facilitated the spread of political propaganda and disinformation on encrypted messaging apps. Based on our research, these strategies are far-reaching and adaptive and include *old school* propaganda tactics like planting fake stories to be picked up by legitimate news organizations, the use of human-powered over automated approaches to spread content, internet shutdowns combined with limitations to the spread of information from legitimate sources, as well as groups that are hired (often indirectly, by a third-party) to spread state propaganda for pay. Another important source of false and misleading information on encrypted messaging apps is false information spread directly by people without intent to mislead. This is particularly insidious on EMAs, given the prevalence of community groups of friends and family members, which inspires trust in the information shared in these spaces.

False and misleading information results in amorphous harms that, while often incalculable, can be disastrous for freedom of expression and democracy writ large. We have identified several troubling harms in our qualitative analysis. First, we heard about disinformation and propaganda on EMAs resulting in the spread of offline – especially ethnic – violence in several countries facing conflict and/or humanitarian crises, such as Ethiopia, Eritrea, and India. Second, the aim of using disinformation to manipulate elections, while pervasive across countries, was especially

highlighted in Indonesia, the Philippines, and Mexico. Finally, disinformation on EMAs, in particular, has caused issues for civil society and fact-checkers who often struggle to enter these spaces. The inability of fact-checkers to consistently access these spaces paired with the increased trust in information coming from known community members such as friends, family, and neighbors makes disinformation on EMAs of particular concern.

While political propaganda and disinformation spread in encrypted spaces limits freedom of expression in many cases, doing away with encryption as a policy solution would arguably do even more harm to the free flow of information and political activism, particularly in authoritarian countries. Though in some countries, such as Morocco and to a certain extent Turkey, we heard that activists are so certain of the governments' far-reaching power that even end-to-end encryption does not offer a safe haven for organizing, many activists we spoke with said that encrypted chat apps were the main, if not the only, avenue for democratic activism when living under an authoritarian regime. In particular, Burmese activists said they received all their trustworthy news from Telegram and engaged in citizen fact-checking to temper misinformation. An activist in Egypt told us that activism is now online-only, and Signal is integral to that effort. Thus, while EMAs are being mobilized to prop up authoritarian governments, they remain indispensable to human rights defenders and activists in their fight against those very same regimes. Furthermore, journalists have grown skilled in using EMAs to promote their own writings, partially freeing them from regime oversight or even censorship. This development provides hope for the emergence of many voices and alternative arguments even in countries that are defined by centrally coordinated top-down communication, such as Egypt, for example.

It is precisely this difficulty that necessitates careful approaches to interventions that take harms seriously and create effective means of addressing them, while not sacrificing the right to private communication. It stands to reason that governments are concerned with encrypted messaging apps as venues that allow for many egregious activities – i.e., the coordination and/or dissemination of content related to political violence, terrorism, and the spread of child sexual abuse material (CSAM) – as well as other problematic content such as false and misleading information, harassment, spam, and more. But we do not argue that the spread of false and misleading information should ever justify the end of encryption - quite the opposite. In a post-Roe America, encryption attains a new level of importance and emerges as a litmus test for the protection of human rights of people who want to get abortions. Maintaining encryption while at the same time providing means to be able to prosecute perpetrators of terrorism or propagators of child sexual abuse material becomes imperative.

Against this backdrop, and to safeguard individuals' identities, their safety, and their free speech, content moderation in a classical sense must be reframed and rethought within the context of EMAs. Content moderation as we see it now on popular social media platforms such as Facebook, Twitter and TikTok emphasizes a balancing act between tech and government oversight. This cannot and should not be companies' modus operandi when it comes to encrypted spaces, as it will likely open the door for added surveillance and a complete erosion of privacy and free speech so valued by our interviewees around the world, but also to marginalized groups that have struggled to carve out safe, private spaces. Instead, moderation on encrypted messaging apps should emphasize user agency, as well as strengthen and motivate intentionally designed reporting mechanisms modeled after the experiences of its most vulnerable stakeholders and communicate effectively and transparently disclosure policies. Unfortunately, legislation that proposes to break encryption by forcing companies to create "backdoors" for law enforcement agencies and government bodies such as the EARN IT Act in the United States and the CSAM proposal in the European Union threatens to upend efforts that

have been made in that direction and does away with already-shrinking protected spaces for human rights defenders, journalists, and marginalized individuals.

As important as technology companies' work on protecting encryption while shielding vulnerable populations from harm is to raise the profile of content moderation as a venerable activity itself, one that goes hand in hand with online participation. The idiosyncrasies of encrypted chat and messaging platforms put the onus of moderation directly on those who find themselves confronted with problematic content. This means that learning how to intervene becomes an important civic responsibility - one that can be taught and learned. This 'moderation literacy' forms a crucial lever, combined with the need to raise self-responsibility of online populations. Ongoing research on volunteer and community moderation can teach a great deal how digital spaces can maintain community norms and boot nefarious actors and content. Such research also shows a path in which moderation constitutes a collective activity of people who engage online, understood as social corrective action and civic labor.

Prepared by Azza El-Masri, Zelly Martin, and Martin J. Riedl