Graphika

BRIEFING FOR THE UNITED STATES HOUSE OF REPRESENTATIVES COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY

Investigations and Oversight Subcommittee Hearing on Online Imposters and Disinformation

Statement of Camille Francois, Chief Innovation Officer, Graphika, and Affiliate at the Berkman Klein Center for Internet & Society, and Ben Nimmo, Director of Investigations, Graphika

Washington, DC
September 26, 2019

The threat represented by the proliferation of information operations designed to deceive and manipulate users on social media demands a unified, forceful response by the whole of society.

The problem is nuanced and complex. There is enormous diversity in the types of actors, techniques, and impacts across the many campaigns our team has analyzed over the past few years.

To date, Facebook, Twitter, and Google alone have detected and taken down information operations emanating from at least 25 different countries, some of them designed to control domestic politics, others targeting geopolitical rivals. The countries named as points of origin for these operations, most of which were not directly attributed to state actors, include Russia, Iran, Bangladesh, Venezuela, Spain, China, Saudi Arabia, Ecuador, United Arab Emirates, Egypt, Myanmar, Iraq, Ukraine, Israel, Thailand, the Philippines, Honduras, India, Indonesia, Pakistan, the United Kingdom, Romania, Moldova, Macedonia, and Kosovo.

Some of these operations were directly coordinated by state actors. The best known are Russia and Iran, but China,[1] Honduras, and others also belong on the list.[2] Campaigns in Spain[3] and India[4] were linked to political parties; others were run by shadowy marketing companies, mercenary firms that execute influence operations on behalf of their clients, as witnessed in Israel,[5] Egypt, and the United Arab Emirates.[6] Some appeared linked to individual media outlets, as in the case of Kremlin-sponsored outlet Sputnik,[7] small groups of activists, as in the United Kingdom,[8] or even to specific individuals with political agendas, as in the Philippines during the recent senatorial election.[9]

---

[1] https://newsroom.fb.com/news/2019/08/removing-cib-china/
[2] https://newsroom.fb.com/news/2019/07/removing-cib-thailand-russia-ukraine-honduras/
[3] https://newsroom.fb.com/news/2019/09/removing-coordinated-inauthentic-behavior-in-spain/
[4] https://newsroom.fb.com/news/2019/04/cib-and-spam-from-india-pakistan/
[5] https://newsroom.fb.com/news/2019/05/removing-coordinated-inauthentic-behavior-from-israel/
[6] https://newsroom.fb.com/news/2019/08/cib-uae-egypt-saudi-arabia/
[7] https://newsroom.fb.com/news/2019/01/removing-cib-from-russia/
[8] https://newsroom.fb.com/news/2019/03/removing-cib-uk-and-romania/
[9] https://newsroom.fb.com/news/2019/03/cib-from-the-philippines/

The global impact of deliberate manipulations of political conversations on social media is difficult to quantify, but a few data points can help us grasp orders of magnitude.[10] Since October 2018, Twitter has published 25,084 accounts associated with information operations in ten different countries and has confirmed that tens of thousands more low-grade spam accounts were also involved in similar behavior.[11] The full archive of information operation[12] posts shared by Twitter encompasses over 65 million tweets, spanning more than five years of posting.

Over the past two years, Facebook has announced taking down 12,085 accounts, pages, groups, and Instagram accounts for engaging in what it calls "coordinated inauthentic behavior." Just over 40 million other accounts followed one or more of these assets. The least-followed operation gathered under 1,000 followers, while the most followed, run by commercial companies in the United Arab Emirates and Egypt, gathered over 13 million.[13] Reach measured in number of followers here is a very imperfect proxy for impact, though: *who* those followers are and *how* they are engaged often matter more than *how many* they are.

For the individual user, 50 million tweets or 40 million followers are almost too many to visualize, but for the platforms themselves, with hundreds of millions of active users, they represent only a fraction of daily activity. For the operators, meanwhile, what matters most is often a small group of deliberately chosen targets: a protest community, a politically influential group, or even an individual journalist who might unwittingly spread the desired narratives and alter their behavior based on anything from an artificially boosted trend to the release of hacked[14]materials.[15] The impact of these operations, from ruined reputations, to gaming the journalistic agenda, to election dynamics, are very real. The increase in information operations since 2016, and the range of actors carrying them out, should be ample evidence of the effectiveness of these methods.

It is critical to understand that these types of operations long predate 2016. Iran's known operations targeted US audiences with fake social media profiles as early as 2013. Russia's Internet Research Agency began attacking domestic opposition on Russian-language channels as far back as 2010 and further developed these methods in the 2014 Ukrainian conflict while ramping up US involvement in the same year. As early as 2012, the campaign of Mexican presidential challenger (and eventual president) Enrique Pena Nieto was accused of benefiting from large-scale amplification by Twitter bots (automated accounts), nicknamed Penabots. This problem has been with us for a while.

Unfortunately, it took until the Russian interference in the 2016 US election to force us toward a collective reckoning. In 2016, the major platforms, law enforcement, and democratic institutions sleepwalked through the Russian assault on US democratic processes, and those in the open-source community who raised the alarm were, at best, ignored. As just one example, Russian operators ran a Twitter account that claimed to be the unofficial outlet of the Republican Party in Tennessee and registered it to a Russian mobile phone number, yet the account survived three

[10] These figures are based on the platforms' public announcements, made intermittently through their blog posts. As a result, they represent voluntary disclosures on incidents that have been investigated to date, and are therefore not fully representative of the scale of the problem.
[11] https://about.twitter.com/en_us/values/elections-integrity.html#data
[12] For a searchable archive, see our efforts to make this data more available to the public on www.io-archive.org
[13] https://newsroom.fb.com/news/2019/08/cib-uae-egypt-saudi-arabia/
[14] On hack-to-leaks operations, see our work on *False Leaks* at CYBERWARCON 2018: https://www.youtube.com/watch?v=P8iXN8j4gMk
[15] https://www.nature.com/articles/d41586-019-02235-x

complaints to Twitter from the actual Tennessee Republican Party.[16] Even when disinformation became a national security issue for American democracy, we collectively failed to properly recognize and address it. It's fair to say that back then, most platforms were unaware of the scale or seriousness of this type of activity, did not have applicable rules against it, and weren't actively looking to protect their users from it.

We've come a long way since then. The main platforms, but also investigative journalists, government actors, and a network of skilled researchers are now actively looking for, investigating, and coordinating to take down influence operations across a wide range of online environments. They have begun working with external researchers, both to expose more operations and to explain what they have found. These green shoots are promising and should be commended. But unfortunately, we are not the only ones making progress.

There are now more actors perpetrating information operations, and primary adversaries are better resourced and more sophisticated every day. Facebook confirmed in July 2018 that the Internet Research Agency's operators were taking ever more effective steps to mask their presence, including using internet phone numbers to register accounts and proxy servers in third countries to mask their origins. They also paid third parties to run ads on their behalf.[17] A separate Russian operation, exposed in June 2019 and suspected of being run by an intelligence service, went a step further by creating hundreds of blogs and social media accounts to post forged documents and divisive content and then abandoned most of the accounts after they had posted just once.[18] A government-linked operation in China used large numbers of hijacked and repurposed accounts to spread its message.[19] An operation emerging from Iraq used stolen official personal identification documents in an attempt to avoid systems in place to detect false accounts.[20] And as the mainstream platforms crack down on information operations, we also see operators invest in alternative and more marginalized platforms to run operations in more permissive environments. We tend to focus on the major platforms, but as their efforts to combat bad actors become more effective, the problem is migrating to smaller platforms that lack the capabilities and, sometimes, the will to fight back. We have documented one Russian operation alone that worked across more than 30 different platforms to spread false narratives.

So how do we tackle this issue together? Disinformation and information operations present a multi-faceted problem requiring technical, methodological, and policy solutions borrowed from disciplines as diverse as cybersecurity, data science, and consumer protection and privacy law. We need to understand all vectors critical to a disinformation campaign's impact: *Manipulative actors*, *deceptive behaviors*, and *harmful content.* These three vectors are what we call the "ABC's of disinformation."[21] Content elements, like "deep fakes," get the most public attention, but their impact depends on more hidden but critical matters of *how* that content is being disseminated and *who* is hiding behind a campaign.

---

[16] https://www.buzzfeednews.com/article/kevincollier/twitter-was-warned-repeatedly-about-this-fake-account-run#.tjENWBAQlv

[17] https://newsroom.fb.com/news/2018/07/removing-bad-actors-on-facebook/

[18] https://medium.com/dfrlab/top-takes-suspected-russian-intelligence-operation-39212367d2f0

[19] https://www.aspi.org.au/report/tweeting-through-great-firewall

[20] https://newsroom.fb.com/news/2019/09/removing-coordinated-inauthentic-behavior-from-iraq-and-ukraine

[21] See attached paper: Francois, Camille. "Actors, Behaviors, Content: A Disinformation ABC. Highlighting Three Vectors of Viral Deception to Guide Industry & Regulatory Responses," A working paper of the Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression: https://www.ivir.nl/twg/.

Science and technology have a crucial role to play in tackling this problem. The sheer volume of information on these platforms, and the speed with which it is shared, require new methods for campaign detection that can scale beyond our current capabilities. As our opponents become more effective at concealing their identities, we need to continuously innovate by creating forensic approaches that will be both accurate and difficult to undermine. And for us to make real, measurable progress on these fronts, we need to address the thorny but essential problem of data availability.

The task at hand is to design a system that guarantees user security and privacy while ensuring that academic researchers, cybersecurity professionals, and human rights investigators can access the data they need to unlock our understanding of these threats and harness innovative ways to tackle the issue. Today, we're very far from such a system.

Let me illustrate: The most well-understood campaign ever, for which the most data to date has been made available by all platforms, is the Russian campaign targeting the American public around the 2016 election. The trove of data released by US platforms and institutions has enabled superb academic work by our colleagues.[22] The Graphika team, along with our colleagues at the Oxford Internet Institute, spent seven months investigating additional, non-public data on behalf of the Senate Intelligence Committee.[23]

Our confidence in the completeness of this picture is false. There remain critical data blind spots. For instance, while platforms released a trove of data regarding the Internet Research Agency's public posts on social media, little to nothing has been shared regarding the GRU's campaigns, when in reality the GRU is the better funded and more persistent actor. It is also inherently more threatening, given its advanced hacking capabilities and readiness to leak apparently compromising material. Similarly, we know that Russian operators used private messaging to target and cultivate relationships with activists, campaign staffers, and journalists, but there is no data available anywhere to indicate a sense of scale and no public records to learn from to determine how best to immunize future targets against these types of threats. Finally, we know that the Russian operators designed their messages to be inflammatory and sometimes overtly hateful: how many of these posts have been moderated by platforms long before anyone cared about Russian trolls? Are we missing a large chunk of content from the public record? We believe that these data blind spots undermine our preparation for the threats ahead. We can, and must, do better.

Data availability is a major part of the solution. Another is ensuring that a community of scholars and practitioners exists to leverage it. At present, the study of these kinds of information operations on social media is a nascent discipline. We need help to turn it into a comprehensive and cooperative field that brings together experts who span the social and data sciences under a common framework, and with common goals.

---

[22] See for instance: Stewart, Leo G., Ahmer Arif, and Kate Starbird. "Examining trolls and polarization with a retweet network." In Proc. ACM WSDM, Workshop on Misinformation and Misbehavior Mining on the Web. 2018. ; Boatwright, Brandon C., Darren L. Linvill, and Patrick L. Warren. "Troll factories: The internet research agency and state-sponsored agenda building." Resource Centre on Media Freedom in Europe (2018) ; Benkler, Yochai, Robert Faris, and Hal Roberts. Network propaganda: Manipulation, disinformation, and radicalization in American politics. Oxford University Press, 2018. (Chapter 8: *Are the Russians coming?*).

[23] Howard, Philip N., Bharath Ganesh, Dimitra Liotsiou, John Kelly, and Camille François. The IRA, social media and political polarization in the United States, 2012-2018. University of Oxford, 2018.

In summary, the emerging field of interdisciplinary scholars and practitioners dedicated to ensuring the integrity of online conversations needs support, funding, and shared infrastructure that allow effective and collaborative innovation. We need this field to keep maturing and growing, to blossom into a community of ethnographers, historians, data scientists, cognitive psychologists, computer scientists, political scientists, sociologists, and many others. The diversity of this community will enable us to address our biggest challenges with a variety of informative perspectives. In this way, we will continue to collaboratively build upon a robust set of interdisciplinary methods, scientific rigor, and shared ethical principles.