POMEPS
STUDIES
43

Digital Activism and Authoritarian Adaptation in the Middle East

August 2021

Project on Middle East Political Science
Stanford | CDDRL | Program on Arab Reform and Democracy
Stanford | Global Digital Policy Incubator | Cyber Policy Center

Contents

Preface 3
Larry Diamond and Eileen Donahoe, Stanford University

Digital Activism and Authoritarian Adaptation in the Middle East 4
Marc Lynch, Project on Middle East Political Science

Binary Threat: How Governments' Cyber Laws and Practice
Undermine Human Rights in the MENA Region 8
Ahmed Shaheed, University of Essex, & Benjamin Greenacre, City University of New York

The Implementation of Digital Surveillance Infrastructures in the Gulf 16
James Shires, Leiden University

The web (in)security of MENA civil society and media 22
Alexei Abrahams, University of Toronto

Beyond Liberation Technology? The Recent Uses of Social Media by Pro-Democracy Activists 29
Joshua A. Tucker, New York University

Chinese Digital Authoritarianism and Its Global Impact 35
Xiao Qiang, University of California at Berkeley

Transnational Digital Repression in the MENA Region. 41
Marwa Fatafta, Access Now

Social media manipulation in the MENA: Inauthenticity, Inequality, and Insecurity. 48
Andrew Leber, Harvard University and Alexei Abrahams, University of Toronto

Tracking Adversaries and First Responding to Disinfo Ops:
The Evolution of Deception and Manipulation Tactics on Gulf Twitter 56
Marc Owen Jones, Hamid Bin Khalifa University

Follow the Money for Better Digital Rights in the Arab Region 63
Afef Abrougui, Independent Consultant and Researcher and Mohamad Najem, Executive Director, SMEX

Digital Orientalism: #SaveSheikhJarrah and Arabic Content Moderation 69
Mahsa Alimardani and Mona Elswah, Oxford Internet Institute, University of Oxford

Official Foreign Influence Operations: International Broadcasters in the Arab Online Sphere 76
Alexandra A. Siegel, University of Colorado - Boulder

Russian Digital Influence Operations in Turkey 2015-2020 83
Akin Unver, Ozyegin University and Ahmet Kurnaz, Çanakkale Onsekiz Mart University

Middle East Influence Operations: Observations Across Social Media Takedowns 91
Renée DiResta, Stanford University; Josh A. Goldstein, Stanford University; and Shelby Grossman, Stanford University

Changing Sources: Social Media Activity During Civil War. 103
Anita Gohdes, Hertie School, Berlin and Zachary C. Steinert Threlkeld, University of California, Los Angeles

The Project on Middle East Political Science

The Project on Middle East Political Science (POMEPS) is a collaborative network that aims to increase the impact of political scientists specializing in the study of the Middle East in the public sphere and in the academic community. POMEPS, directed by Marc Lynch, is based at the Institute for Middle East Studies at the George Washington University and is supported by Carnegie Corporation of New York and the Henry Luce Foundation. For more information, see http://www.pomeps.org.

The Center on Democracy, Development and the Rule of Law

The Center on Democracy, Development and the Rule of Law (CDDRL) at Stanford University is an interdisciplinary center for research on development in all of its dimensions: political, economic, social, and legal, and the ways in which these different dimensions interact with one another. CDDRL does not simply seek to study democracy, development and the rule of law; we think these phenomena embed critical values that we believe in and want to promote. CDDRL was launched in 2002. Its first director was Coit Blacker, followed by Stephen Krasner, Michael McFaul, Larry Diamond, and Francis Fukuyama. For more information, see https://cddrl.fsi.stanford.edu/.

The Global Digital Policy Incubator at the Stanford Cyber Policy Center

The mission of the Global Digital Policy Incubator at the Stanford Cyber Policy Center is to inspire policy and governance innovations that reinforce democratic values, universal human rights, and the rule of law in the digital realm. Its purpose is to serve as a collaboration hub for the development of norms, guidelines, and laws that enhance freedom, security, and trust in the global digital ecosystem. The Global Digital Policy Incubator provides a vehicle for global multi-stakeholder collaboration between technologists, governments, private sector companies, diplomats, international organizations, academics, and civil society in a shared purpose: to develop norms and policies that enhance security, promote economic development, and reinforce respect for human rights in or our global trans-border digital ecosystem. For more information, see https://cyber.fsi.stanford.edu/gdpi.

Preface

The essays in this collection are the fruit of a collaboration between the Project on Middle East Political Science (POMEPS) and two Stanford University research programs: the Global Digital Policy Incubator (GDPI, based at the Cyber Policy Center) and the Program on Arab Reform and Democracy (ARD, based at the Center on Democracy, Development, and the Rule of Law). As leaders of the latter two programs, we would like to express our appreciation to Marc Lynch and his colleagues in POMEPS, especially Tessa Talebi, and to our own program colleagues, Hesham Sallam of ARD and Tracy Navichoque of GDPI. Most of all we want to thank the authors for their papers, their insights, and their patient commitment to this project, which was delayed by the onset of the COVID pandemic.

This project is coming to fruition at an increasingly troubling time for freedom and democracy, both in the Arab world and globally. Over the last decade, the bright political hopes of the 2010-2011 Arab uprisings have given way to political polarization, violence, coups, and in a few cases, state breakdown. As we publish these essays, an authoritarian executive coup is unfolding in the one Arab country that was able to move from protest to democracy—Tunisia. The wealthy and technologically sophisticated Gulf states have not only set the regional standard for digital surveillance, repression, and control, they have also lent generous political, financial, and technical support and encouragement to their embattled or unstable authoritarian peers in the region. And they have intensified repression of their own citizens through digital technologies of censorship and information control. China is the world's leader in materializing George Orwell's nightmarish vision of omniscient totalitarian monitoring of individuals and pervasive state control and manipulation of information. But Gulf states like the UAE and Saudi Arabia are coming up fast in these capacities. As our papers make clear, the trend toward digital authoritarianism in the Middle East also draws crucial support from outside the region, not only through the technology exports and cross-border information operations of authoritarian mega-powers like China and Russia, but through the promiscuous transfer of spyware and other digital surveillance tools and expertise by private companies based in Western democracies and especially notably of late, in Israel.

Yet our essays caution against overly gloomy or deterministic forecasts. As in other regions, civil society activists adapt and innovate to use and widen available spaces. As we see from the recent protests in Lebanon, Algeria, and Sudan, from the substantial boycott of Iran's recent presidential "election", and from new and ongoing forms of activism elsewhere in the region, as well as from multiple rounds of the Arab Barometer, people in the Middle East still aspire for the same basic political ideals that drove the Arab uprisings: dignity, voice, accountability, and self-determination. Thus, the public sphere remains contested, even embattled, in cyberspace, as it periodically does in the streets as well. And just as authoritarian powers and amoral corporations have aided Middle Eastern states in their ambitions to extend control, there remains considerable scope for the world's democracies to help tip the balance toward freedom and accountability through financial and technical assistance and diplomatic support for the region's creative, courageous, and tenacious netizens. They are not going away.

Larry Diamond and Eileen Donahoe
Stanford University
July 29, 2021

Digital Activism and Authoritarian Adaptation in the Middle East

Marc Lynch, Project on Middle East Political Science

Social media platforms and digital technologies played a decisive role in political mobilization, before, during and after the 2011 Arab uprisings; inspiring academic and popular discussions of the internet as a "liberation technology" inevitably undermining the foundations of authoritarian states.[1] But it is no longer 2011. The naïve assumption that "the internet" necessarily would serve as a liberation technology has been dislodged by overwhelming evidence to the contrary, as authoritarians have discovered creative ways to capitalize on digital technologies for repression and control.[2] The ubiquity of online infrastructures has facilitated new forms of digital authoritarianism, through surveillance, manipulation, disinformation, and highly targeted repression.[3] The use of such tools by state and non-state actors now presents a major challenge not only to activists in authoritarian contexts but to democracies.

To explore these issues, the Project on Middle East Political Science partnered with Stanford University's Center for Democracy, Development and the Rule of Law and its Global Digital Policy Incubator for an innovative two week online seminar. This workshop built upon more than a decade of our collaboration on issues related to the internet and politics in the Middle East, beginning in 2011 with a series of workshops in the "Blogs and Bullets" project supported by the United States Institute for Peace and the PeaceTech Lab.[4] This new collaboration brought together more than a dozen scholars and practitioners with deep experience in digital policy and activism, some focused on the Middle East and others offering a global and comparative perspective. POMEPS STUDIES 43 collects essays from that workshop, shaped by two weeks of public and private discussion.

The rapid development of digital infrastructures forces analysis to move beyond last decade's debates about online versus offline, social media vs broadcast media, liberation vs repression.[5] Digital tools that were once novel have become ubiquitous, with internet use now nearly universal across most of the region and with no easy separation between the virtual and the real. Media ecosystems, as Ethan Zuckerman reminds us, cannot in any useful way be understood as a set of discrete online and offline platforms.[6] What could it mean to say "Twitter caused X" or "Clubhouse could lead to Y" when those platforms are fully integrated into dense, richly interwoven communication networks? Broadcast media stream over mobile devices and maintain popular websites and social media feeds, while videos and ideas from social media cross smoothly and seamlessly into satellite television programs and print publications.

This ubiquity and integration has significant analytical implications. Questions which dominated the literature in the early 2010s such as "does social media empower political protest" no longer make sense, when social media are so fully integrated into media ecosystems and the broader political realm. It is better to think in terms of socially mediated public spheres, where conversations, information and sentiments fluidly travel through multiple platforms, and in terms of discrete mechanisms by which particular communication flows might shape attitudes, behavior or outcomes.[7] Television broadcasts with vast audiences may appear to fit a traditional model of one-to-many broadcasting, but it is the complex, rapid simultaneous discussions of those programs – the sharing of clips through social networks, the retweeting or commenting on key moments, the formation of clusters of attitudes around their contents – which constitutes the public sphere.

Social media tools are not only, or even primarily, used for political mobilization. They have transformed every dimension of social, cultural and political life, as they have woven themselves into the fabric of everyday life. Particularly in a region dominated by the young, it is almost impossible today to even remember a time when

people did not receive news, share opinions, or experience popular culture through social media. Public facing apps such as Facebook, Twitter and YouTube dominated the scene during the 2011 uprisings. Since then, new apps regularly appear and rapidly gain currency: WhatsApp and Telegram, with their combinations of encrypted one to one communication and large scale groups for in-network semi-public sharing and conversation; Clubhouse, with its live audio discussions; TikTok, whose playful videos can slide easily into political statement. Some of these new apps, as Joshua Tucker argues in this collection, fill similar functions as the older generation of social media for political mobilization. But others present substantive differences: in-network communications, such as those on WhatsApp or Telegram, may be invisible to researchers but central to lived experience. Their encryption possibilities may also provide an unwarranted sense of safety to their users. In all cases, seemingly apolitical apps can quickly take on political roles for creative activists – which, in turn, increases the incentive of autocratic regimes to control and surveil them.

The ubiquity of socially mediated communication intersects uneasily with the pervasive and potent memories of the revolutionary moment of 2011 which still inspire activists and frighten autocrats. The more that activists and autocrats alike recognized the potency of digital communications, the more they sought to use it and bend it towards their own purposes. Middle Eastern regimes, focused primarily on preventing challenges to their own survival, came to view social media as a major potential threat and as such, invested heavily in ways to control, surveil and manipulate online activity. The push towards digital authoritarianism took many forms: colonization of the online public through manipulation, inauthentic activity, and influence operations; surveillance of the online public through big data analysis, spyware, and tracking apps; silencing of the online public through deplatforming, content moderation and targeted repression of influential voices; transnationalization of repression through information operations and surveillance abroad; and the normalization and legalization of architectures of digital control. The MENA public sphere today is shaped by the ubiquity, then, of both socially mediated digital communication and transnational digital authoritarianism.

The trend towards digital authoritarianism, both globally and regionally, has progressed through a combination of technology, policy and law. While regimes may have in the past aspired to these kinds of control and surveillance, the material possibilities have only become available relatively recently. The reach and scope of surveillance technology is now breathtaking, with online life tailor-made to offer visibility into the political and private lives of its users. As the papers in this collection document, regional governments have been enthusiastic consumers of the most advanced surveillance tech, such as those revealed in investigations of the NSO Group and Pegasus. Most famously, the assassination of Saudi dissident journalist Jamal Khashoggi appears to have been driven by intercepted online communications; but that is only the tip of the iceberg in terms of state efforts to spy on potential and real opponents. James Shires points out the importance of the physical infrastructures of digital surveillance in the Gulf, including the locating of digital clouds in Saudi Arabia and the regional offices of platforms such as Facebook and Twitter in the UAE. The normalization of relations between the UAE and Israel has accelerated the already robust market for Israeli surveillance tech in the Gulf. Some states, such as Iran, may aspire to a Chinese style of state bounded and controlled internet. For all their adaptability and courage, many civil society actors have left themselves at the mercy of this surveillance technology, as Alexei Abrahams shows in his original research on Palestinian civil society.

The ultimate goal of this surveillance infrastructure might be seen in the Chinese model explored by Xiao Qiang, where comprehensive surveillance becomes a societal norm, while highly sophisticated artificial intelligence assesses massive quantities of data to identify threats, trends and opportunities for state action. Such a goal may be out of reach for many poor, low capacity Arab states which can barely manage the basics of governance. But for the wealthy, high capacity states of the Gulf, it is not only a goal but increasingly a reality – one accelerated by

the adoption of COVID-tracking apps which increasingly normalize ubiquitous surveillance and state visibility into every aspect of citizen lives. The UAE has gone the farthest in this direction, layering sophisticated digital surveillance into its already pervasive authoritarianism and state domination of society. Other regimes in the region would surely prefer to follow suit.

Digital authoritarianism has technological, policy, and legal dimensions. As Ahmed Shaheed and Benjamin Greenacre point out in meticulous detail, the region's autocrats have sought not only to engage in surveillance and manipulation, but also to craft a permissive regulatory and normative framework. Norms and ethics rarely outweigh power politics or economic opportunities, as Mohammed Najem and Afef Abroughi discuss, in the digital realm or elsewhere. Multiple exposes and sustained criticism by human rights organizations have done little to halt the transnational diffusion of intrusive digital surveillance tech. It is difficult to shame the shameless or impose reputational costs in an atmosphere of impunity.

Activists continue to adapt and evolve, of course. Joshua Tucker shows in his global overview that activists have adapted new platforms and technologies to achieve familiar aims such as coordination, broadcasting, and overcoming barriers to collective action. This summer's Palestinian mobilization against efforts by Israeli settlers to seize homes in the Sheikh Jarrah neighborhood of East Jerusalem offers a vivid example of the continuing power of social media to break through information blockades, generate local and international support, and reshape political realities. The very power and success of that online mobilization brought its own response, of course, as Israel successfully lobbied social media platforms like Facebook and Instagram to remove large volumes of Palestinian content while nontransparent algorithms "unintentionally" blocked even more. It speaks to the stakes of these online narrative battles that Israel has taken such aggressive measures to police and control them.

As many of the essays note, the active role being played by platforms today is an important new dimension.[8] The traditional assumption of the platforms as essentially neutral brokers, providing a level playing field for the actors to fight it out, no longer holds. There have always been algorithmic biases nudging users towards particular types of content. The new element is active intervention by platform managers, such as Twitter banning Donald Trump or Facebook agreeing to huge volumes of Israeli demands to take down Palestinian content and remove Palestinian users. Through content moderation and takedowns, algorithmic promotion, and selective enforcement, social media platforms such as Facebook, Twitter and Instagram are increasingly taking sides in contentious political struggles. Why they do so remains a matter of debate. Mohammed Najem and Afef Abroughi point to the economic incentives, with social media platforms eager to retain access to lucrative markets and investment capital. Mona Elswah and Mahsa Alimardani instead offer a cultural and political narrative grounded in what they deem an Orientalist policing of Arab and especially Palestinian content.

This digital authoritarianism is increasingly transnational in nature. One dimension of this is that regimes which view threats to their stability from dissidents living abroad can use digital methods to identify and surveil them without regard to national borders. Khashoggi's murder and widely reported threats to other Saudi dissidents abroad have received the most attention here, but other countries such as Morocco, Egypt, Iran and Turkey have carried out similar digital surveillance of their citizens abroad. Marwa Fatafta's essay unpacks the various dimensions of these efforts by MENA states to exercise control beyond their borders, extending their repressive reach across a genuinely transnational public sphere.

A second form of transnational digital politics is the crowded realm of disinformation, manipulation and information warfare.[9] State and non-state actors use a wide range of inauthentic activity to shape narratives, promote particular ideas or politicians, and interfere with events across borders.[10] Hamit Akin Unver and Ahmet Kurnaz argue in their essay for a more comparative approach to online disinformation. Their detailed study of Russian

information operations in Turkey highlights a range of possible behaviors, in line with foreign policy goals. Andrew Leber and Alexei Abrahams warn that it is too simplistic to simply assume that armies of bots are behind digital narrative warfare.[11] Often real people are the key influencers, even if bots serve as amplifiers. Identifying bots and inauthentic activity remains critical, however, as Marc Owen Jones explains – both for policy makers and for academics hoping to learn from big online data. Academic researchers must also be wary, as Zachary Steinert-Threlkeld and Anita Gohdes warn, of how the composition of users changes over the course of war. Their study of geotagged Twitter users from Syria shows that it is not only messages and content which change, but the actual users – which could have significant implications for how we interpret trends in discourse.

What about policy interventions against disinformation and manipulation? Renee DiResta and Shelby Grossman do a deep dive into takedowns reported by Facebook and Twitter, showing both what they have chosen to do and its limits. MENA regimes have been, perhaps unsurprisingly, at the forefront of state-sponsored inauthentic activity, with networks associated with Iran, Saudi Arabia and Egypt among the most frequently targeted for removal. Alexandra Siegel shows in her contribution that platforms labeling media outlets as state-controlled can significantly impact their reach. Other authors point to possibilities for human rights naming and shaming campaigns and the mobilization of international law and norms to limit the transnational reach of digital authoritarianism.

Taken together, the essays in this collection offer a rigorous, empirically rich and theoretically sophisticated snapshot of an embattled digital public sphere in the Middle East. The extent to which repressive states have been empowered as a consequence of digitization can seem overwhelming. But, as Tucker and others remind us in this collection, activists have consistently found ways to exploit new apps, and found creative ways to leverage online platforms for mobilization and information sharing. Social media platforms themselves may find incentives to change their approach, should they face sufficient reputational costs. And overly intrusive state censorship and surveillance could trigger its own backlash, pushing typically apolitical citizens into opposition. We expect to continue exploring these constantly evolving digital politics in our ongoing project.

Endnotes

1 Larry Diamond and Marc F. Plattner, eds. *Liberation Technology: Social Media and the Struggle for Democracy*.(Baltimore: Johns Hopkins University Press, 2012); Sean Aday, Henry Farrell, Deen Freelon, Marc Lynch and John Sides. *Blogs and Bullets II: The Impact of New Media on the Arab Uprisings. Washington DC: United States Institute for Peace Peaceworks (2012); Marc Lynch, After Egypt: The Promise and Limitations of the Online Challenge to the Authoritarian Arab State. Perspectives on Politics 9, no.2 (2011): 301-18;* Nils B. Weidmann and Espen G. Rød, *The Internet and political protest in autocracies* (Oxford, UK: Oxford University Press, 2019).

2 Larry Diamond, "Rebooting Democracy." *Journal of Democracy 32, no.2 (2021): 179-183; Joshua A. Tucker, Yannis Theocharis, Margaret E. Roberts, and Pablo Barberá. "From liberation to turmoil: Social media and democracy." Journal of Democracy 28, no. 4 (2017): 46-59.*

3 Ron Deibert, "Authoritarianism Goes Global: Cyberspace Under Siege," *Journal of Democracy 26, no. 3 (July 13, 2015): 64–78.*

4 Sean Aday, Henry Farrell, Marc Lynch and John Sides *Blogs and Bullets: New Media and Contentious Politics*. United States Institute of Peace (2010); Sean Aday, Henry Farrell, Deen Freelon, Marc Lynch and John Sides *Blogs and Bullets II: The Impact of New Media on the Arab Uprisings*. United States Institute for Peace Peaceworks (2012); Marc Lynch, Deen Freelon and Sean Aday, Blo*gs and Bullets III: Syria's Socially Mediated Civil War*, US Institute for Peace (2014); *Blogs and Bullets IV: How Social Media Undermines Transitions to Democracy*. With Deen Freelon and Sean Aday. Washington, D.C.: PeaceTech Lab (2016).

5 James Shires, *The Politics of Cybersecurity in the Middle East (London, UK: Hurst/Oxford University Press, 2021).*

6 Ethan Zuckerman, "Why Study Media Ecosystems?" *Information, Communication and Society. Online First (2021), DOI: 10.1080/1369118X.2021.1942513*

7 Henry Farell, "The Consequences of the Internet for Politics," *Annual Reviews of Political Scienc*e 15 (2012): 35-52.

8 Jillian C. York, *Silicon Values : The Future of Free Speech Under Surveillance Capitalism* (Verso, 2021)

9 Marc Owen Jones "The Gulf Information War| Propaganda, Fake News, and Fake Trends: The Weaponization of Twitter Bots in the Gulf Crisis". International Journal Of Communication, 13, 27 (2019); Marc Owen Jones, Digital Authoritarianism, Deception, Disinformation and Social Media. (London: Hurst, forthcoming).

10 Samantha Bradshaw and Philip N. Howard. "The global organization of social media disinformation campaigns." *Journal of International Affairs* 71, no.1 (2018): 23-32.

11 Alexi Abrahams and Andrew Leber. "Comparative Approaches to Mis/Disinformation| Electronic Armies or Cyber Knights? The Sources of Pro-Authoritarian Discourse on Middle East Twitter." *International Journal of Communication* 15 (2021): 1173-1199.

Binary Threat:
How Governments' Cyber Laws and Practice Undermine Human Rights in the MENA Region

Ahmed Shaheed, University of Essex, & Benjamin Greenacre, City University of New York[1]

Introduction

As the papers in this collection document, the emancipatory promise of technology is overshadowed by a rising tide of States who co-opt technological advances to enable online and offline repressive measures, a phenomenon otherwise referred to as the rise of 'digital authoritarianism'. This trend has been thrown into high relief in the Middle East and North Africa (MENA) region by geopolitical shifts that have given greater influence to states that are importing and exporting repressive technologies, applications, and governance models. Digital authoritarianism does not involve just the co-option of technology, but also the re-shaping and subversion of international norms to reduce the transactional costs of authoritarian control and suppress the legitimacy of mobilization for greater online and offline freedoms. This paper examines the laws and practices of states, including through Covid-linked state responses, in the MENA region that enable digital authoritarianism and their disjuncture with the human rights obligations of these states. It then shows that despite this regression, the 'cat-and-mouse' contest between digital authoritarianism and digital activism is a fluid one, highlighting opportunities to push back against this authoritarianism.

Legal Frameworks

The right to privacy and freedoms of expression, association and assembly are guaranteed by articles 17, 19, 21 and 22 of the International Covenant on Civil and Political Rights (ICCPR), an instrument ratified by all states of the MENA region bar Oman, Saudi Arabia, and the United Arab Emirates. These rights are also widely considered part of customary international law under the Universal Declaration of Human Rights (UDHR)[2] and therefore binding on all States. These rights are further reiterated by Article 21, 24 and 32 of the Arab Charter on Human Rights. Nonetheless, the past decade has been marked by a proliferation of legislation across the MENA region that restricts and even criminalises legitimate expression, association and assembly and privacy in digital spaces.

So-called 'Cybercrime' laws drafted in the Gulf States,[3] Egypt,[4] Iran[5] and Jordan,[6] fall far short of international standards. Where a State wishes to impose a restriction on freedom of expression, it must, *inter alia,* draft a clear, precise, and unambiguous provision (the principle of legality)[7] that is both necessary and proportionate. The provision must equally be lawful; for example, restrictions can never be used to muzzle "advocacy of democratic tenets and human rights."[8] However, many cybercrime laws characterise legitimate expression as potentially criminal activity. In the UAE, for example, Decree No. 5/2012 is used as a legal basis for the prosecution of individuals who use technology to criticise the government, argue for political reform, or organise unlicensed demonstrations.[9] Jordan's cybercrime bill punishes digital libel and a vague conception of 'hate speech' with up to three years in prison and punitive fines.[10] Many of these laws do not satisfy, in letter or effect, the basic requirements of restrictions on freedom of expression under international law.

Freedom of expression also includes the right to receive information of all kinds, such as political discourse, commentary on public affairs and human rights, and journalism.[11] The right to receive information is, in particular, a bellwether for the health of democratic norms and the enjoyment of human rights, online and offline, within a country. However, internet censorship is fast becoming the norm rather than the exception in the

Middle East and North Africa.[12] The Committee to Protect Journalists lists Saudi Arabia and Iran as two of the most censored countries globally, blocking vast swathes of the internet deemed objectionable under their respective cybercrime legislation, especially regional human rights monitoring organisations.[13] Similarly, Egyptian authorities have blocked access to 513 websites, including news media and prominent human rights organisations under the country's 2018 Cybercrime laws.[14]

Some MENA States use existing criminal laws, sometimes in combination with cybercrime laws, to limit expression online. These often result in severe or increased penalties, which incentivise people to self-censor on specific topics. Kuwait,[15] Jordan,[16] Egypt[17], Iran[18] and Saudi Arabia[19] have all prosecuted human rights defenders for expression online under anti-blasphemy laws. These provisions can result in heavy prison sentences and, in Saudi Arabia and Iran, the death penalty. Regularly used to stifle political and cultural dialogue online and offline,[20] blasphemy laws protect religious institutions and symbols from insult or offence at the expense of the freedom of religion or belief and freedom of expression of actual rights-holders. Iran,[21] Saudi Arabia[22], Jordan,[23]and Kuwait[24] also use overly broad defamation laws to prosecute individuals that criticise the government or spotlight corruption on online platforms, threatening them with prison sentences and punitive fines. International law is clear that, while protecting the rights and reputations of others is a valid reason to restrict freedom of expression, restrictions cannot legally impede political debate[25] or shield political figures or institutions from criticism.[26]

At the extreme, States are using military, anti-terrorism, and national security laws to undermine freedom of expression and association online. As with anti-blasphemy laws, these laws often carry the most severe of penalties and are, in some jurisdictions, routinely meted out by military tribunals, which the UN Human Rights Committee has noted may violate the right to a fair trial.[27] In particular, Israel regulates digital expression and association in the Occupied Palestinian Territories (OPT) through the incitement provisions of Israeli military law.[28] During 2015-2018 Israel reportedly prosecuted over 500 Palestinians under these provisions, including children and journalists sharing news on Facebook.[29] Anti-terrorism provisions are similarly levied against human rights defenders in areas under civil jurisdiction of Israel,[30] as well as in Kuwait[31] and Jordan.[32] In many cases, the conviction of one individual as a terrorist may provide grounds for the prosecution by association of all individuals in their online networks.

Invasive Government Laws

These new digital regulatory environments also actively facilitate infringements on the right to privacy. The majority of countries in the region, including Egypt, UAE, Bahrain and Qatar, have some legal provisions for data protection that should further the enjoyment of the right to privacy.[33] However, the vast majority of these laws provide insufficient protections against unauthorised processing of subjects' sensitive data and contain numerous and significant exemptions allowing State security services to carry out invasive domestic surveillance.[34] Governments' failure to properly regulate digital data is problematic. While legitimate security exigencies upon the State call for limited infringement of rights, weak digital data regulation is directly linked to an increase in arbitrary and unlawful infringements of the right to privacy.[35] Problematic in of themselves, infringements of the right to privacy also further stymy freedom of expression, notably as they encourage regimes of self-censorship.[36]

State Practice

Unsurprisingly, many governmental practices and activities enabled by the above legislative frameworks also violate the international human rights obligations, norms and standards. For the purposes of this essay, two problem categories are identified: digital surveillance practices and digital interference practices.

Digital surveillance is now widespread throughout the MENA region. As States seek to build their surveillance capabilities, they have purchased complex surveillance equipment, spyware packages (software which covertly

gathers data from your computing devices and transmits them to a third party without your consent) and 0-day exploits (vulnerabilities in software or hardware that are not publicly known) from private technology companies. At first, this was predominately from international companies such as Nokia Siemens, Palantir, Fin Fisher, BAE Systems and Hacking Team, however notable local enterprises have since emerged, such as the Israeli 'NSO Group' and Emirati 'DarkMatter'.[37] Such firms have, alongside criminal mercenary hacker groups, significantly augmented the capacity of governments including Bahrain, Saudi Arabia, Egypt, Morocco, Algeria, UAE, Sudan and others to mount complex, targeted or large-scale digital surveillance operations over the past decade, often with little regard as to the human rights impact.[38]

Even if a government plans to use such services in accordance with international law, they may still facilitate violations of a variety of rights by supporting such a model of surveillance. Complex state-funded spyware, once discovered, can be reverse engineered and repurposed for malicious use.[39] The more States fund the proliferation of spyware, the more this will occur. Further, many forms of monitoring rely on software vulnerabilities that anyone can exploit for as long as they remain unpatched. States relying on such vulnerabilities are therefore enabling other unknown actors to access the same data.[40]

Where a State does not meet its obligations under international law concerning digital surveillance practices, surveillance can escalate directly to grave violations of human rights, including arbitrary detention, torture, and even extrajudicial killing of individuals a government considers hostile.[41] There is no clearer example of this than the case of journalist Jamal Khashoggi, who was tortured and dismembered by the Saudi Arabian government after the contents of his encrypted chats with other dissidents were compromised by NSO Group's 'Pegasus' spyware.[42]

Digital interference, by contrast, includes practical measures taken by governments to block, limit, or distort access to information within their jurisdiction. Examples regularly seen in the MENA region include internet shutdowns and throttling (the reduction of internet speed to render services or content effectively unusable) and the manipulation of online narratives through bot networks and troll farms. In 2020 the MENA region had the second largest number of internet shutdowns, with Yemen and Jordan having among the highest number of all counties globally.[43] Interference with access to the internet most obviously undermines the freedom to impart and receive information. However, given the internet's prominence in all aspects of our life, especially during the COVID-19 pandemic, internet shutdowns imperil the enjoyment of numerous rights, including by interfering with the ability of persons to manifest their religion or belief in a community, to work or receive an education, to participate in politics and to receive healthcare information. Moreover, internet shutdowns can carry an implicit threat to the right to life, often preceding atrocities.[44] In particular, Iran has shut down internet services during protests immediately before disproportionate lethal responses by security forces.[45] As a restriction on speech, internet blackouts are inherently disproportionate due to their blanket effect and, therefore, an unlawful restriction on digital expression.[46]

Social media bots and troll farms, by comparison, are means for altering the normative and discursive environment in a subtler fashion. They are often a form of organised disinformation, where States "systematically and simultaneously suppress other sources while promoting their own false narratives".[47] Although these activities may not restrict online activity in of themselves, they are used to dispirit and demoralise activists, either through targeted harassment or through an overwhelming hijacking of the online narrative, against which authentic activity struggles to compete. In Saudi Arabia, an infamous troll farm allegedly run by former Saudi advisor Saud Al-Qahtani, harassed critics, including Jamal Khashoggi, and ensured that the Saudi-authority favoured narrative was the only one that would 'trend' on social media platforms.[48] Similar efforts have been a staple of Iran's approach to manipulating online discursive environments.[49] Practices of online disinformation can explicitly violate freedom of opinion and expression and may interfere with freedom of thought, where they undermine our mental autonomy.

In some instances, partnerships with private companies have allowed MENA State actors to interfere with the online efforts of individuals to exchange information, views and opinions simply by requesting that companies remove content from their influential social media platforms. At the behest of the governments of Tunisia, Syria, Palestine, and Egypt, digital media platforms including Facebook, Twitter, YouTube, and Instagram have censored and disabled the accounts of activists, journalists, and citizens critical of their governments.[50]An agreement between the Government of Israel and Facebook has resulted in 95% of requests submitted by the government for content removal being upheld, including for alleged incitement. Given the contention surrounding the use of Israel's incitement laws, rights monitors are concerned that moderation disproportionately and discriminatorily targets Palestinians.[51] Accusations of discriminatory censoring of Palestinians by Facebook has escalated rapidly in May 2021, as Palestinians began documenting their evictions from Sheikh Jarrah, clashes with Israeli police at Al-Aqsa Mosque and the renewed hostilities between Israel and Hamas.[52] Governments imposing discursive norms on social media platforms via their moderation practices are particularly concerning, as, unlike coordinated inauthentic behaviour, it is much harder to discern and object to an absence of information.

COVID-19 Related Intrusions

The global COVID-19 pandemic has exacerbated the adverse human rights impacts of law and policy affecting digital spaces. Governments' technology responses to the COVID-19 pandemic – usually involving some variant of a 'track and trace' app – have profoundly impacted the right to privacy. In this regard, Amnesty International declared Bahrain and Kuwait's COVID tracking apps 'among the most dangerous in the world,' categorising them as "highly invasive surveillance tools," which "go far beyond what is justified in efforts to tackle COVID-19."[53] Some apps, such as Qatar's ETHERAZ, have been made mandatory for all residents and visitors and are required for entering many different shared spaces, including mosques, entertainment, social venues, public transport, parks, schools and childcare establishments.[54] The app's security protocols and data retention policy have been a cause for concern, both due to the possibility for hackers to access users' sensitive personal information and for the government's ability to use the data for various purposes beyond preventing the transmission of COVID. Like Qatar, most governments in the region have not developed COVID tracing apps with a 'clear and limited' purpose, with 'data protection by design and default.'[55] Where access to public places is predicated upon the user's consent to be monitored by such apps, governments are forcing rights holders to choose between their right to privacy and freedom of movement, freedom of religion or belief, the right to education and or the right to work. Where there are no such limitations, governments are still at best putting the sensitive data of their citizens at risk and at worst expanding the surveillance state under the guise of medical imperative.

The COVID pandemic has further been used to justify restrictions on free speech under the guise of combating medical misinformation. In Algeria, for example, authorities have arrested journalists, bloggers and others who contradict or criticise the government's COVID narrative online.[56] Similarly, Egypt, Israel, Palestine, Jordan, Iraq, Bahrain, Iran and Tunisia have used COVID-19 emergency measures or other laws to arrest, detain, prosecute or fine persons expressing opposition to the government's pandemic response – or even criticising the government on unrelated issues.[57] Some Gulf governments have exceptionally lifted restrictions on access to voice over IP and encrypted chat apps during the pandemic.[58] These gestures remain insufficient, however, as long as lawful access remains transient and popular apps such as WhatsApp and Skype remain inaccessible. The on-going ban of WhatsApp in these countries may in fact harm efforts to combat medical misinformation as many international organisations including the WHO distribute COVID information via WhatsApp chat bots.[59]

Empowering People: Creative Counter Responses

Despite rampant surveillance and censorship over the past decade, social media and digital technology continue

to amplify the voices of human rights activists and promote accountability of both State actors and private enterprise. In 2019, Rahaf Mohammad Mutlaq Al-Qunun avoided being deported to Saudi Arabia, where she feared execution for apostasy, primarily due to her documenting on Twitter the attempts of Thai authorities to force her repatriation.[60] On May 12, 2021, due to the documentation and outcry of rights monitors, Facebook was forced to acknowledge its failure to appropriately moderate discussions on its platform surrounding Israeli police operations at the al-Aqsa mosque.[61]

Rather than cowing civil society, 'digital authoritarianism' has instead become a double-edged sword for governments, simultaneously restricting digital activism and encouraging activists to innovate new approaches to counteract repression. Virtual private networks (VPNs), encrypted communication and peer-to-peer networking and file-sharing have enabled human rights activists to bypass State censor regimes. Meanwhile, the deluge of primary sources across social media, combined with the dissemination of information and expertise across global networks of activists, has given birth to entirely new methodologies of accountability. One of the most popular is 'Open-Source Intelligence' or OSINT.

OSINT is intelligence that is generated, cross-referenced, and verified using publicly available (generally digital) information. The organization 'Bellingcat,' a prominent pioneer of OSINT methods, has used this methodology to uncover the Syrian regime's use of chemical weapons against its civilian population[62] (later confirmed by the Organisation for the Prohibition of Chemical Weapons[63]), the killing of Iraqi protestors by security forces,[64] and Saudi Arabia's long-standing attempts to conduct malicious cyber operations.[65] The methodology's success has resulted in many human rights NGOs[66] and even UN Fact-Finding Missions[67] adopting it to document human rights violations and rebut State disinformation campaigns.

Another response has been the growing application of artificial intelligence (AI) and machine learning (ML) algorithms to data gathered through digital networks. The '*Ceasefire*' platform, the result of a collaboration between Ceasefire, Minority Rights Group, and the University of Essex, combines crowd-sourced reporting of violations with AI & ML processing of social media feeds to pinpoint human rights violations as they occur.[68] The platform has quickly become an important tool not only for its original creators, but for a wide variety of civil society actors to monitor and respond to human rights violations. Piloted in Iraq in 2017, the use of the platform was subsequently extended to the wider Middle East and North Africa region, proving to be a vital source information on human rights violations.[69]

Conclusion

If, in the Middle East & North Africa, the "route to democratisation is a digital one,"[70] then it should also be concluded that, increasingly, so too is the route to authoritarianism. Many States in the region have studied the lessons of Iran's Green Movement and the wider 'Arab Spring.' In response, Governments have not only enacted broad programmes of legislation that restrict and often criminalise many legitimate forms of expression and association online, but developed counter-practices of unprecedented, invasive surveillance, censorship and denial of service.

Further, governments increasingly seek to co-opt social media platforms and tools to advance behind-the-scenes strategies aimed at surreptitiously moderating content, shrinking civil society space, and undermining the promise of technology for accountability. In this context, the global COVID-19 pandemic has served as an excuse for governments, both in the region and globally, to exercise and consolidate repressive power, online and offline.

Nevertheless, despite evident power imbalances, human rights activists have adapted. The practice of adopting and sharing operational security protocols, as well as the use of encryption technology and censor circumvention methods, have provided civil society with the means to maintain digital communication networks. Through these networks, activists coordinate, rapidly exchange information and

advocate to a global audience –– activities that would otherwise be compromised by digital authoritarianism. These networks have additionally enabled civil society to develop entirely new, effective tools to spotlight violations of human rights and combat governments' disinformation campaigns, including by carrying out investigations of a calibre once the exclusive reserve of State intelligence agencies.

Moreover, global efforts to assert the relevance of human rights law in digital spaces, develop more stringent standards for the protection of privacy and data online, resist and circumvent 'sovereign internets,' and defend intermediary immunity can impede the onslaught of digital authoritarianism. The investment made by repressive states, such as China (see Xiao Qiang's essay in this collection), to control multilateral institutions underscore the important role hegemonic norms play in this contestation, as do recent efforts made by technology companies to demonstrate greater compliance with human rights norms. However, the demand that human rights law be applied to the design, development, use and evaluation of emerging technologies,[71] while critically important, does not guarantee that human rights will be protected in the digital age. Leadership by democratic states through investing in technological and institutional infrastructure, disseminating best practice, developing inclusive governance models, and boosting the capacity of human rights defenders, will be crucial to tilt the balance in favour of freedom. Digital authoritarianism in the MENA region is not a localized phenomenon or challenge; it is representative of wider global trends and requires a multi-layer, global response.

Endnotes

1 Ahmed Shaheed is Deputy Director of the Human Rights, Big Data and Technology Project at the University of Essex and Benjamin Greenacre is Senior Researcher at Freedom of Religion or Belief Project at the Ralph Bunche Institute at the Graduate Center at City University of New York.

2 Hurst Hannum, "The UDHR in National and International Law," *Health and Human Rights* 3, no. 2 (1998): 145-150, accessed 24 July 2021, https://www.hhrjournal.org/archives/volume-3-issue-2/.

3 Joyce Hakmeh, "Cybercrime Legislation in GCC Countries," Chatham House, accessed 21 July 2021, https://www.chathamhouse.org/2018/07/cybercrime-legislation-gcc-countries/freedom-expression-online-under-gcc-cybercrime-laws.

4 International Labour Organisation, "Egypt: Law No.175 of 2018 Regarding Anti-Cyber and Information Technology Crimes," NATLEX Database of National Labour, Social Security and Related Human Rights Legislation, accessed 21 July 2021, http://ilo.org/dyn/natlex/natlex4.detail?p_lang=en&p_isn=108464&p_country=EGY&p_count=499.

5 Secretariat of the Working Group for Determining Instances of Criminal Content, Attorney General's Office, Islamic Republic of Iran, "Full Text of the Cybercrime Law," accessed 22 July 2021, https://internet.ir/law.html.

6 Dima Samaro and Emna Sayadi, "Cybercrime Law in Jordan: Pushing Back on New Amendments That Could Harm Free Expression and Violate Privacy," Access Now, accessed 21 July 2021, https://www.accessnow.org/cybercrime-law-in-jordan-pushing-back-on-new-amendments-that-could-harm-free-expression-and-violate-privacy/.

7 See, for example, United Nations, Human Rights Committee, *Leonardus Johannes Maria de Groot v. The Netherlands*, Communication No. 578/1994, UN Doc CCPR/C54/D/578/1994 (14 July 1995), accessed 24 July 2021, http://hrlibrary.umn.edu/undocs/html/dec578.htm.

8 United Nations, Human Rights Committee, *General Comment No. 34, Article 19: Freedom of Expression*, UN Doc. CCPR/C/GC/34 (12 September 2011), para. 23, accessed 24 July 2021, https://www2.ohchr.org/english/bodies/hrc/docs/gc34.pdf.

9 Human Rights Watch, "UAE: Cybercrimes Decree Attacks Free Speech," accessed 21 July 2021, https://www.hrw.org/news/2012/11/28/uae-cybercrimes-decree-attacks-free-speech.

10 Sevan Araz, "Jordan adopts Sweeping Cybersecurity Legislation," Middle East Institute, accessed 21 July 2021, https://www.mei.edu/publications/jordan-adopts-sweeping-cybersecurity-legislation.

11 See: United Nations, Human Rights Committee, *Primo Jose Essono Mika Miha v. Equatorial Guinea,* Communication No. 414/1990, UN Doc. CCPR/C/51/D/414/1990 (10 August 1994), accessed 24 July 2021, http://hrlibrary.umn.edu/undocs/html/vws414.htm; Human Rights Committee, *Patrick Coleman v. Australia,* Communication No. 1157/2003, UN Doc. CCPR/C/87/D/1157/2003 (10 August 2006), accessed 24 July 2021, http://hrlibrary.umn.edu/undocs/1157-2003.html; Human Rights Committee, *Vladimir Velichkin v. Belarus,* Communication No. 1022/2001, UN Doc. CCPR/C/85/D/1022/2001 (20 October 2005), accessed 24 July 2021, http://hrlibrary.umn.edu/undocs/1022-2001.html; and Human Rights Committee, *Rakhim Mavlonov and Shansiy Sa'di v Uzbekistan*, Communication No. 1334/2004, UN Doc. CCPR/C/95/D/1334/2004 (19 March 2009), accessed 24 July, https://juris.ohchr.org/Search/Details/1486.

[12] Helmi Noman, "Internet Censorship and the Intraregional Geopolitical Conflicts in the Middle East and North Africa," Internet Monitor, accessed 21 July 2021 https://thenetmonitor.org/bulletins/internet-censorship-and-the-intraregional-geopolitical-conflicts-in-the-middle-east-and-north-africa.

[13] Committee to Protect Journalists, "Ten Most Censored Countries," accessed 21 July 2021 https://cpj.org/reports/2019/09/10-most-censored-eritrea-north-korea-turkmenistan-journalist/#4

[14] Amnesty International, "Egypt: Series of Draconian Laws 'Legalizes' Unprecedented Repression Six Years Since Fall of Morsi," accessed 21 July 2021, https://www.amnesty.org/en/latest/news/2019/07/egypt-series-of-draconian-laws-legalizes-unprecedented-repression-six-years-since-fall-of-morsi/.

[15] Global Freedom of Expression, Columbia University, "The Case of Hamad Al-Naqi (Kuwait Twitter Blasphemy Case," accessed 21 July 2021, https://globalfreedomofexpression.columbia.edu/cases/the-case-of-hamad-al-naqi-kuwait-twitter-blasphemy-case/.

[16] Amnesty International, "The Killing of Jordanian Journalist: A Deplorable Attack on Freedom of Expression," accessed on 21 July 2021, https://www.amnesty.org/en/latest/news/2016/09/killing-of-jordanian-journalist-a-deplorable-attack-on-freedom-of-expression/.

[17] Mahmoud Mourad, "Egyptian Poet Goes on Trial Accused of Contempt of Islam," Reuters, accessed 22 July 2021, https://www.reuters.com/article/us-egypt-courts-poet-idUSKBN0L121M20150128.

[18] Humanists International, "Iran," accessed 22 July 2021, https://end-blasphemy-laws.org/countries/middle-east-and-north-africa/iran/.

[19] See: Global Freedom of Expression, Columbia University, *Saudi Public Prosecutor v. Raif Badawi*, Jeddah Criminal Court Case No. 242/183/29 (7 May 2014), accessed 22 July 2021, https://globalfreedomofexpression.columbia.edu/cases/saudi-public-prosecutor-v-raif-badawi/

[20] Agnes Callamard, "Freedom of Speech and Offence: Why Blasphemy Laws are Not the Appropriate Response," *Equal Voices*, Issue 18 (June 2006):3-4, accessed 24 July 2021, https://www.article19.org/data/files/pdfs/publications/blasphemy-hate-speech-article.pdf.

[21] Center for Human Rights in Iran, "Detained Editor Who Exposed Corruption Slapped with Additional Charge," accessed 22 July 2021, https://www.iranhumanrights.org/2016/09/yashar-soltani/.

[22] Amnesty International, "Waleed Abu al-Khair, Imprisoned in Saudi Arabia for Defending Human Rights," accessed 22 July 2021, https://www.amnesty.org.uk/saudi-arabia-free-human-rights-lawyer-waleed-abulkhair-abu-al-khair.

[23] Arab Network for Human Rights Information, "Jordan: Three Months in Prison for Prolonging the Tongue in an Opinion Case,' accessed 22 July 2021, http://anhri.net/?p=144952.

[24] Human Rights Watch, "Kuwait: Drop Charges for 'Offending Emir'," accessed 22 July 2021, https://www.hrw.org/news/2013/04/15/kuwait-drop-charges-offending-emir.

[25] Human Rights Committee, *General Comment No. 34,* para. 28.

[26] See: United Nations, Human Rights Committee, *Zeljko Bodrožić v Serbia and Montenegro*, Communication No. 1180/2003, UN Doc. CCPR/C/85/D/1180/2003 (31 October 2005), accessed 24 July 2021, http://hrlibrary.umn.edu/undocs/1180-2003.html; and Human Rights Committee, *Rafael Marques de Morais v. Angola*, Communication No. 1128/2002, UN Doc. CCPR/C/83 /D/1128/2002 (29 March 2005), accessed 24 July 2021, http://hrlibrary.umn.edu/undocs/1128-2002.html.

[27] United Nations, Human Rights Committee, *General Comment No. 32, Article 14: Right to Equality Before Courts and Tribunals and to Fair Trial,* UN Doc. CCPR/C/GC/32 (23 August 2007), paras 13-14, accessed 24 July 2012, https://www.refworld.org/docid/478b2b2f2.html.

[28] See section 251 (b) of Israeli Military Directive No. 1651 of 2009. Available in English from http://www.militarycourtwatch.org/files/server/MO%201651%20.pdf [accessed 22 July 2021].

[29] Anan AbuShanab, "Connection Interrupted: Israel's Control of the Palestinian ICT Infrastructure and Its Impact on Digital Rights," 7amleh-Arab Center for the Advancement of Social Media (December 2018):35, accessed 22 July 2021, https://7amleh.org/wp-content/uploads/2019/01/Report_7amleh_English_final.pdf

[30] *Ibid*, p. 36

[31] Global Freedom of Expression, Columbia University, "The Case of Hamad al-Naqi,", accessed 21 July 2021, https://globalfreedomofexpression.columbia.edu/cases/the-case-of-hamad-al-naqi-kuwait-twitter-blasphemy-case/.

[32] Human Rights Watch, "Country Summary: Jordan," January 2017, pp. 1-2, accessed 22 July 2021, https://www.hrw.org/sites/default/files/jordan_1.pdf

[33] Social Media Exchange, "Data Protection and Privacy Laws in MENA: A Case Study of Covid-19 Contact Tracing Apps," accessed 22 July 2021, https://smex.org/wp-content/uploads/2021/02/210210_JoeyShea_Report_Covid-19ContactTracingApps_EN_Draft5.pdf.

[34] Marwan Fatafta and Dima Samaro, "Exposed and Exploited: Data Protection in the Middle East and North Africa," Access Now, January 2021, accessed 22 July 2021, https://www.accessnow.org/cms/assets/uploads/2021/01/Access-Now-MENA-data-protection-report.pdf.

[35] United Nations, Human Rights Council, *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue,* UN Doc. A/HRC/23/40 (17 April 2013), para. 3, accessed 24 July 2021, https://undocs.org/A/HRC/23/40.

[36] Elizabeth Stoycheff, "Under Surveillance: Examining Facebook's Spiral of Silence Effects in the Wake of NSA Internet Monitoring," *Journalism and Communications Studies Quarterly* 93, no. 2 (June 2016): 296-311, accessed 24 July 2021, https://doi.org/10.1177/1077699016630255; and Alex Marthews and Catherine E. Tucker, "Government Surveillance and Internet Search Behaviour," SSRN Electronic Journal (17 February 2017), accessed 24 July 2021, SSRN: https://ssrn.com/abstract=2412564.

[37] Jon Hoffman, "Espionage and Repression in the Middle East Courtesy of the West,' Open Democracy, accessed 22 July 2021, https://www.opendemocracy.net/en/north-africa-west-asia/espionage-and-repression-middle-east-courtesy-west/.

[38] *Ibid*. See also BlackBerry, "BlackBerry Uncovers Massive Hack-For-Hire Group Targeting Governments, Businesses, Human Rights Groups and Influential Individuals," accessed 22 July 2021, https://www.blackberry.com/us/en/company/newsroom/press-releases/2020/blackberry-uncovers-massive-hack-for-hire-group-targeting-governments-businesses-human-rights-groups-and-influential-individuals.

[39] Andy Greenberg, "A China-Linked Group Repurposed Hacking Team's Stealthy Spyware," Wired, accessed 22 July 2021, https://www.wired.com/story/hacking-team-uefi-tool-spyware/.

[40] Joseph Menn, "Spy Agency Ducks Questions about 'Back Doors' in Tech Products," Reuters, accessed 22 July 2021, https://www.reuters.com/article/us-usa-security-congress-insight-idUSKBN27D1CS.

[41] United Nations, Human Rights Council, *Surveillance and Human Rights: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression,* UN Doc. A/HRC/41/35 (28 May 2019), para. 1, accessed 24 July 2021, https://undocs.org/A/HRC/41/35.

[42] Access Now, "Two Years after Khashoggi's Slaying, No Accountability for Spyware Firm or Saudi Government," accessed 22 July 2021, https://www.accessnow.org/khashoggi-two-years-later/.
[43] Berhan Taye and Access Now Team, "Shattered Dreams and Lost Opportunities: A Year in the Fight to #KeepIOn," Access Now for the #KeepItOn Coalition (March 2021): 5, accessed 22 July 2021, https://www.accessnow.org/cms/assets/uploads/2021/03/KeepItOn-report-on-the-2020-data_Mar-2021_3.pdf.
[44] Access Now, "Primer on Internet Shutdowns and the Law," (November 2016): 12-15, accessed 22 July 2021, https://www.ohchr.org/Documents/Issues/Expression/Telecommunications/AccessPart_I.docx.
[45] Amnesty International, "Iran: Internet Deliberately Shut Down During November 2019 Killings – New Investigation," accessed 22 July 2021, https://www.amnesty.org/en/latest/news/2020/11/iran-internet-deliberately-shut-down-during-november-2019-killings-new-investigation/
[46] United Nations, Human Rights Council, *Disinformation and Freedom of Opinion and Expression: Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan,* UN Doc. A/HRC/47/25 (13 April 2021), para. 51, accessed 24 July 2021, https://undocs.org/A/HRC/47/25.
[47] *Ibid.,* para. 47.
[48] Katie Benner, Mark Mazzetti, Ben Hubbard and Mike Isaac, "Saudis' Image Makers: A Troll Army and a Twitter Insider," The New York Times, accessed 22 July 2021, https://www.nytimes.com/2018/10/20/us/politics/saudi-image-campaign-twitter.html.
[49] Nathaniel Gleicher, "Removing Coordinated Inauthentic Behavior from Iran," Facebook, accessed 22 July 2021, https://about.fb.com/news/2019/01/removing-cib-iran/.
[50] Access Now, "Open Letter to Facebook, Twitter, and YouTube: Stop Silencing Critical Voices from the Middle East and North Africa," accessed 22 July 2021, https://www.accessnow.org/facebook-twitter-youtube-stop-silencing-critical-voices-mena/.
[51] Adalah, "Adalah Fears Facebook's Online Incitement Deal with Israel will Selectively Target Palestinian Citizens," accessed 22 July 2021, https://www.adalah.org/en/content/view/8948.
[52] Kassem Mnejja and Marwa Fatafta, "Sheikh Jarrah: Facebook and Twitter Systematically Silencing Protests, Deleting Evidence," Access Now, accessed 22 July 2021, https://www.accessnow.org/sheikh-jarrah-facebook-and-twitter-systematically-silencing-protests-deleting-evidence/.
[53] Amnesty International, "Bahrain, Kuwait and Norway Contact Tracing Apps among Most Dangerous for Privacy," accessed 22 July 2021, https://www.amnesty.org/en/latest/news/2020/06/bahrain-kuwait-norway-contact-tracing-apps-danger-for-privacy/.
[54] Social Media Exchange, 'Data Protection," pp. 10-11.
[55] "Data protection by design and default', which is essential to the safeguarding of the right to privacy, means privacy protection must be integrated from the outset when designing a system and that privacy respecting settings must be enabled by default. See United Nations, Human Rights Council, *The Right to Privacy in the Digital Age: Report of the United Nations High Commissioner for Human Rights,* UN Doc. A/HRC/39/29 (3 August 2018): para. 31, accessed 24 July 2021, https://undocs.org/A/HRC/39/29.
[56] Amnesty International, "Algeria 2020," accessed 22 July 2021, https://www.amnesty.org/en/countries/middle-east-and-north-africa/algeria/report-algeria/.
[57] Human Rights Watch, "Covid-19 Triggers Wave of Free Speech Abuse," accessed 22 July 2021, https://features.hrw.org/features/features/covid/index.html.
[58] Human Rights Watch, "COVID-19: Unblock Voice Over IP Platforms in Gulf," accessed 22 July 2021, https://www.hrw.org/news/2020/04/07/covid-19-unblock-voice-over-ip-platforms-gulf.
[59] World Health Organization, "WHO Health Alert Brings COVID-19 Facts to Billions via WhatsApp," accessed 22 July 2021, https://www.who.int/news-room/feature-stories/detail/who-health-alert-brings-covid-19-facts-to-billions-via-whatsapp.
[60] Anais Moine, "Tout Juste Reconnue Réfugiée, Rahaf Mohammed Témoigne de Son Bonheur," Aufeminin, accessed 22 July 2021, https://www.aufeminin.com/news-societe/declaree-refugiee-rahaf-mohammed-al-qunun-confie-son-bonheur-s2993886.html.
[61] Jon Porter, "Instagram Blames 'Enforcement Error' for Removal of Posts about Al-Aqsa Mosque," The Verge, accessed 22 July 2021, https://www.theverge.com/2021/5/13/22433861/instagram-al-aqsa-mosque-posts-takedown-error-facebook-moderation.
[62] Eliot Higgins, "The Open Source Hunt for Syria's Favourite Sarin Bomb," Bellingcat, accessed 22 July 2021, https://www.bellingcat.com/news/2020/04/21/the-open-source-hunt-for-syrias-favourite-sarin-bomb/.
[63] Organization for the Prohibition of Chemical Weapons, "OPCW Releases Second Report by Investigation and Identification Team," accessed 22 July 2021, https://www.opcw.org/media-centre/news/2021/04/opcw-releases-second-report-investigation-and-identification-team.
[64] Nick Waters, "Iraqi Protesters are Being Killed By "Less Lethal" Tear-Gas Rounds," accessed 22 July 2021, https://www.bellingcat.com/news/mena/2019/11/12/iraqi-protesters-are-being-killed-by-less-lethal-tear-gas-rounds/.
[65] Bellingcat, "Lord Of the Flies: An Open-Source Investigation into Saud Al-Qahtani," accessed 22 July 2021, https://www.bellingcat.com/news/mena/2019/06/26/lord-of-the-flies-an-open-source-investigation-into-saud-al-qahtani/.
[66] See for example, Sam Dubberley, "The Digital Verification Corps: Amnesty International's Volunteers for the Age of Social Media," Amnesty International, accessed 22 July 2021, https://citizenevidence.org/2019/12/06/the-digital-verification-corps-amnesty-internationals-volunteers-for-the-age-of-social-media/. See also Joshua Lyons, "Documenting Violations of International Humanitarian Law from Space: A Critical Review of Geospatial Analysis of Satellite Imagery During Armed Conflicts in Gaza (2009), Georgia (2008), and Sri Lanka (2009)," *International Review of the Red Cross* 94, no 886 (Summer 2012) 741, accessed 24 July 2021, https://www.icrc.org/en/doc/assets/files/review/2012/irrc-886-lyons.pdf.
[67] United Nations, Human Rights Council, *Situation of Human Rights in Yemen, including Violations and Abuses since September 2014: Report of the Detailed Findings of the Group of Eminent International and Regional Experts on Yemen*, UN Doc. A/HRC/42/CRP.1 (3 September 2019), para. 34, accessed 24 July 2021, https://www.securitycouncilreport.org/atf/cf/%7B65BFCF9B-6D27-4E9C-8CD3-CF6E4FF96FF9%7D/A_HRC_42_CRP_1.pdf.
[68] See Ayman Alhelbawy, Mark Lattimer, Udo Kruschwitz, Chris Fox and Massimo Poesio, "An NLP-Powered Human Rights Monitoring Platform," *Expert Systems with Applications* 153 no. 113365 (September 2020), accessed 24 July 2021, https://doi.org/10.1016/j.eswa.2020.113365.
[69] *Ibid*, p. 12.
[70] Philip N. Howard, *The Digital Origins of Dictatorship and Democracy: Information, Technology and Political Islam,* New York: Oxford University Press (2010), p. 201, accessed 24 July 2021, DOI:10.1093/acprof:oso/9780199736416.001.0001 [].
[71] United Nations, Human Rights Council, *The Right to Privacy in the Digital Age: Resolution adopted by the Human Rights Council on 26 September 2019*, UN Doc. A/HRC/RES/42/15 (7 October 2019), accessed 24 July 2021, https://digitallibrary.un.org/record/3837297?ln=en.

The Implementation of Digital Surveillance Infrastructures in the Gulf

James Shires, Leiden University

Introduction

Authoritarian adaptation to political uncertainties associated with the ubiquity of digital communications comes in many forms. Some responses are familiar to scholars of political violence across technological eras: vague national security threats, securitized public spaces, repressive action against protests, and the intimidation, imprisonment, and torture of activists, journalists, and other critical or dissenting voices.[1] Other responses are themselves enabled by digitalization, covered by other papers in this collection, including social media manipulation and transnational information operations. This paper highlights a less visible form of authoritarian adaptation through the constellation of various public and private actors involved in the procurement, installation, and maintenance of digital surveillance infrastructures. It argues that the implementation of such infrastructures – the micro-level expert routines and practices in their design and construction – represents a key site of power and contest overlooked by existing treatments.

This paper draws on my more detailed analysis elsewhere of the politics of cybersecurity in the Middle East.[2] It focuses on digital surveillance infrastructures in the states of the Gulf Cooperation Council (GCC), for two reasons. First, the GCC states are the most "successful" examples of authoritarian adaptation to digital activism in the Middle East.[3] These states largely avoided the violence and chaos that emerged from suppression of the "Arab Spring" movements in 2011 - although GCC military and security assistance to Bahrain and simmering tensions in eastern Saudi Arabia are important exceptions.[4] Domestically, these states are especially alive to the perceived dangers of social media and heavily invested in ensuring revolutionary politics do not rise up again.[5] Regionally, they have sought to influence the course of devastating conflicts in Libya, Syria, and Yemen online and offline, with varied results.[6]

Second, the Gulf states are extremely well-connected, both in terms of their embrace of digital technologies (internet penetration rates are consistently the highest in the region and compare favorably worldwide),[7] and because their openness to global capital has made them sites of significant digital expertise in e-government, energy, health, and other critical sectors.[8] These states have sought to use their reputation as leaders in digital innovation both to underpin long-term efforts at economic diversification and to deflect political criticism. Such claims were called into question after the 2017 Gulf crisis, as both Qatar and the quartet of blockading states sought to undermine the other's international legitimacy.[9] Other regional expert relationships have emerged openly after the 2020 Abraham Accords, with a prominent Israeli role in Gulf digital technology sectors mooted almost before the ink was dry.[10] But flows of expertise are not just inward: digital communications networks in the Gulf have expanded to accompany aggressive foreign policies, as well as providing broader soft power and economic links to the horn of Africa and the Maghreb.[11] As such, although digital surveillance infrastructures are everywhere, they are especially notable in the GCC states: domestically resilient, regionally influential and globally integrated.

The infrastructural tendencies of digital surveillance

No state is content with unreachable communications transiting their territory. The implementation of digital surveillance technologies must therefore be considered in the context of broader information controls, as an explicit policy goal of both authoritarian and democratic states.[12] Of course, such controls vary between states, as well as the

extent to which they are subject to appropriate checks and balances.[13]

Information controls are not necessarily constraining. Many states seek to steer and encourage certain kinds of online activity – nationalistic sentiment, for example - in ways that retain economic advantages and bolster public support for the regime or preferred allies.[14] States can also reshape their information environment indirectly, convincing citizens that pushing the boundaries of acceptable content online is not a good idea. Such "chilling effects" manifest in many forms, from the more physical and bodily forms of intimidation and violence noted above, to the promotion of dominant media narratives and certain forms of cultural and artistic production.[15]

Digital surveillance technologies play a central role in these dynamics. As discussed by Abroughi and Najem in their paper, some of the most high-profile instances of digital surveillance in the Gulf come from various forms of "spyware", designed to hack into specific devices and send data back to the operator.[16] These "targeted" surveillance technologies are generally contrasted, in both academic and policy literature, with "mass" surveillance technologies that enable the near real-time analysis of internet communications at a national scale through deep packet inspection (DPI).[17] Unlike spyware, DPI surveillance can act as the basis for fine-grained internet censorship, by blocking websites or specific applications rather than resorting to blanket internet shutdowns and shunning social media platforms.[18] DPI technologies also enable more subtle forms of interference, such as throttling to slow access speeds or forcing protocols to downgrade to older security standards (see Alexei Abrahams in this collection).[19]

However, a strong distinction between targeted and mass surveillance technologies obscures what I call the infrastructural tendencies of digital surveillance. This is a double move, describing both the impetus for even highly targeted surveillance technologies to be incorporated into wider surveillance architectures, *and* the incorporation of these architectures in turn into the ever-expanding digital infrastructures of states and societies themselves. We can see the first part of this move in the spyware examples discussed by Abroughi and Najem. These spyware companies build and maintain highly complex back-end infrastructures to test, deploy, and receive data from spyware - often constructing separate ones for each task.[20] They can also combine this back-end (otherwise known as "command and control" or C2) infrastructure with DPI monitoring to provide an additional vector for infection. As such, the distinction between targeted and mass surveillance is better described as a spectrum from independent to infrastructural, highlighting how such technologies function within a wider set of capabilities.

The second part of this move draws on the extensive theorization of infrastructures in Science and Technology Studies (STS).[21] Ensmenger notes that "technologies become infrastructure only after they are perfected to the point of being routine."[22] From this perspective, digital surveillance technologies become more infrastructural the more they are accepted as a standard feature of digital societies. Both mass and targeted surveillance technologies function within a wider culture of surveillance that underpins the economic model of social media and, increasingly, our internet-connected lives more generally.[23] This approach to digital surveillance technologies as societal infrastructures helps avoid problematic but popular paradigms of "dual-use", which seek to distinguish – for example – between the use of DPI for increasing efficiency, traffic management, and advertising on one hand, and its use for censorship on the other. Infrastructurally speaking, these are two sides of the same coin. New urban megaprojects in the Gulf, such as Saudi Arabia's NEOM, and the rapid development of "smart city" initiatives elsewhere, provide the ideal ground for this highly infrastructural approach to digital surveillance. To take this insight further, the following section moves from the routinization of digital surveillance infrastructures overall, to the expert routines and practices crucial to their implementation.

Expert practices and surveillance infrastructures

Governments *as a whole* do not implement digital surveillance infrastructures. This task falls to specific actors: communications ministries, telecoms regulatory bodies, intelligence and security organizations, sector-specific or nationwide information technology authorities, and (especially) national cybersecurity institutions.[24] These government actors respond to formal legal requirements or policy directions as well as informal instructions from central figures or more powerful bodies. This delegation of surveillance authority serves several purposes, from cultivating specific technological capabilities, depoliticizing controversial decisions, or even favoring bureaucratic allies in a competitive space.

However, government actors often do not possess the right network position, the necessary equipment, or right kind of technological expertise for digital surveillance, and so they bring in a range of private sector partners. National telecoms companies and internet service providers (ISPs) occupy a central position in this ecosystem, as they often evolve from earlier public-sector entities and retain close elite connections. Telecoms companies and ISPs are mandated by governments to enable access to data centers and crucial nodes for internet traffic, such as connection points to undersea cables (telecoms companies are also often in the consortia that plan and build such cables), as well as responding to specific requests – and sometimes benefitting economically from doing so.[25] A whistleblower speaking to the *Guardian* in 2020 claimed that Saudi mobile telecoms companies had submitted "millions" of tracking requests through a global telecoms protocol (SS7) to determine the location of phones registered in Saudi Arabia while they were roaming in the US.[26] In Iran, these companies have even used their nodal position to re-route internet traffic by exploiting inbuilt protocol characteristics.[27] Their commercial partners provide DPI or similar products, data analysis, or specific capabilities such as geolocation. These partner companies usually originate either in the security and defense sectors (with strong relationships to intelligence agencies or their commercial intermediaries), or in companies offering broader traffic management and performance functions. These companies are often multinational, providing similar solutions across the region and worldwide.

The implementation of digital surveillance infrastructures thus depends not only on technological capacity, but on the norms and practices of relatively small, transnational, expert communities straddling public and private sectors. However, such norms and practices are complex, messy, and at times unpredictable, and in the remainder of this section, I identify four main sources of friction in these expert communities.

First, perhaps the most obvious source of friction is geopolitical. Surveillance solutions are increasingly sold to the Gulf and other attractive economies by rival states or blocs. They can be useful lubricant for diplomatic overtures, as with French/UAE sales of surveillance technologies to Egypt after the seizure of power by President Abdel Fatah Al-Sisi in 2013.[28] But international markets are not easily negotiated. The starkest example is the rift between US and Chinese digital investment, often painted as a choice between Huawei and Cisco routers and other networking equipment (see Xiao Qiang in this volume).[29] In 2020, the US warned the UAE that reliance on Huawei would "risk rupturing [their] long-term strategic relationship", including communication between their respective militaries.[30] My conversations with Huawei employees in the region suggest they are very aware of these dynamics, seeing negotiations with government clients as open competition between superpower surveillance architectures (and these interlocutors repeatedly mentioned the Snowden disclosures to underline US, as well as Chinese, risks). But the US and China are not the only actors in this congested market. An anonymous French surveillance specialist explained to a journalist that the alternative to their sales was "handing control to the Chinese or the Israelis… We tell ourselves we are doing it in the interests of our country."[31] Russian, UK, and German companies and technologies are also regularly in the mix.[32] In this way, the national affiliation

of private providers – both company headquarters and contractor citizenship – complicates authoritarian moves to procure the most effective forms of information control.

Second, installation and maintenance also introduce friction. Updating and recalibrating large-scale surveillance technologies is complicated but essential, not only because the internet itself is constantly changing, but because anti-censorship apps such as Signal and Psiphon also seek to circumvent existing controls. For example, Signal was blocked in December 2016 by the Egyptian and UAE authorities.[33] In response, the creators of Signal worked with activists to update the application to rely on "domain fronting": using encrypted connections to a popular domain, in this case owned by Google, to act as a proxy for Signal messages. Google and Amazon subsequently decided to prevent domain fronting (interestingly, because it presented a "cybersecurity" risk), but by doing so made it easier for authoritarian governments to block Signal and other apps.[34] More generally, retaining qualified and motivated experts – often at significant expense from contractors – is dependent on wider calculations of cost and efficiency and can introduce unintended unreliability into supposedly powerful surveillance solutions.

Third, contracts with international suppliers can come with significant export control requirements, as well as attention from human rights advocates, NGOs, and journalists. The most sustained examination of this space has been conducted by CitizenLab, who discovered in 2011 that Canadian company Netsweeper provided DPI-based filtering technology to the UAE, Qatar, and Kuwait (with the addition of Bahrain in 2016).[35] They then detected that US company Blue Coat had DPI-based filtering and surveillance devices present in all GCC countries other than Oman by January 2013,[36] while McAfee's Smartfilter was identified in the UAE and Saudi Arabia later in 2013.[37] These reports – among many others - comprehensively demonstrate the use of commercial filtering technologies by Gulf governments, most likely through telecoms companies. Despite increasing regulatory attention, export licenses for international suppliers are generally granted for diplomatically convenient sales, facilitated by the wide range of possible "dual-use" reasons for purchase discussed above.[38] Accompanying maintenance obligations mean that such licenses remain crucial for continuing functionality (in the US, through Technical Assistance Agreements or similar mechanisms). Several international suppliers have adopted human rights rhetoric and recommendations – such as ethics committees to review sales – in order to smooth this process, but with little noticeable effect on the wider market.[39]

Fourth, and finally, some states have sought to avoid these international hurdles by investing in domestic equivalents. The most well-known example is the UAE company DarkMatter, which has been reported to implement both targeted and mass surveillance solutions.[40] DarkMatter also applied to become an internet certificate authority, which would have given it greater leverage at an infrastructural level (permission was withdrawn by Mozilla in 2019).[41] However, shifting the locus of surveillance introduces different kinds of friction at the individual rather than corporate level. Key experts freely move between public and private sectors, between countries, and command high salaries. Media reporting indicates that these individuals sometimes reject offers they deem to be morally suspect, as well as reporting misuse or overreach to governments of their home states (for whom they were often originally employed).[42] In some cases, these experts even perform what I call a "moral manoeuvre": altering the implementation of surveillance technologies on the ground, unbeknownst to their clients, to mitigate more extreme requests for surveillance.[43] Overall, the limited size of the expert community, and its transnational connections and dependencies, introduce value judgements and commercial tensions that are key sources of friction in the implementation of digital surveillance infrastructures.

Conclusion

This exploration of digital surveillance infrastructures in the Gulf has demonstrated that implementation matters. Government attempts to reshape norms and practices

of digital surveillance in the Gulf states must navigate technological, social, and political tensions within key expert communities. The choices made by such experts, in specific bureaucratic and commercial contexts, determine how these states enact information controls amid geopolitical, economic, and moral disagreements, thereby contributing to the broader trajectory of authoritarian adaptation in the Middle East. In the future, although relevant expert communities are likely to remain transnational, the increasing territorialization of online activities, including national social media applications, competition over cable connections, and localization requirements for data storage and cloud computing, will probably divide surveillance infrastructures further along global and regional political fault lines. To return to the definition of infrastructure introduced earlier, this paper has approached the authoritarian desire to perfect digital surveillance infrastructures skeptically, suggesting that the routine work of their implementation and maintenance – the possibility of their failure and thus their visibility – will remain an important determinant of authoritarian adaptation in an increasingly digital world.

Endnotes

[1] See e.g. Charles Tilly, *The Politics of Collective Violence* (Cambridge; New York: Cambridge University Press, 2003); Marc Lynch, *The Arab Uprising: The Unfinished Revolutions of the New Middle East* (New York: PublicAffairs, 2013).

[2] James Shires, *The Politics of Cybersecurity in the Middle East* (London, UK: Hurst/Oxford University Press, 2021).

[3] See e.g. Robert Uniacke, "Authoritarianism in the Information Age: State Branding, Depoliticizing and 'de-Civilizing' of Online Civil Society in Saudi Arabia and the United Arab Emirates," *British Journal of Middle Eastern Studies*, March 21, 2020, 1–21.

[4] Toby Matthiesen, *Sectarian Gulf: Bahrain, Saudi Arabia and the Arab Spring That Wasn't* (Stanford, California: Stanford University Press, 2013).

[5] Alexei Abrahams, "Regional Authoritarians Target the Twittersphere," MERIP, December 17, 2019.

[6] Christopher Davidson, *Shadow Wars: The Secret Struggle for the Middle East* (Oneworld Publications, 2016).

[7] Some internet statistics put Qatar and the UAE at over 100% internet penetration (internetworldstats.com). The ITU, using a more conservative methodology, still puts these states at over 99%. Another notable difference is Oman, which has 76.8% internet penetration from internetworldstats and 95.2% from the ITU.

[8] James Shires, "Enacting Expertise: Ritual and Risk in Cybersecurity," *Politics and Governance* 6, no. 2 (2018): 31–40; James Shires and Joyce Hakmeh, "Is the GCC Cyber Resilient?" (London: Chatham House Royal Institute for International Affairs, March 2020).

[9] James Shires, "Disinformation in the Gulf," in *Cyber War & Cyber Peace in the Middle East: Digital Conflict in the Cradle of Civilization*, ed. Michael Sexton and Eliza Campbell (Middle East Institute, 2020), 93–107.

[10] Reuters Staff, "UAE, Israeli Cyber Chiefs Discuss Joining Forces to Combat Common Threats," *Reuters*, September 24, 2020, https://www.reuters.com/article/us-israel-gulf-emirates-cyber-idUSKCN26F2UK.

[11] Adam Hanieh, *Money, Markets, and Monarchies: The Gulf Cooperation Council and the Political Economy of the Contemporary Middle East* (New York: Cambridge University Press, 2018).

[12] Seva Gunitsky, "The Great Online Convergence: Digital Authoritarianism Comes to Democracies," War on the Rocks, February 19, 2020.

[13] Ron Deibert, "Authoritarianism Goes Global: Cyberspace Under Siege," *Journal of Democracy* 26, no. 3 (July 13, 2015): 64–78; Ronald J. Deibert, *Reset: Reclaiming the Internet for Civil Society* (House of Anansi Press Ltd, Canada, 2020).

[14] Andrew Leber and Alexei Abrahams, "A Storm of Tweets: Social Media Manipulation During the Gulf Crisis," *Review of Middle East Studies* 53, no. 2 (December 2019): 241–58.

[15] Adel Iskander. "Re(membering) Culture and Heritage: Egypt's Latest Political Turf War," POMEPS-GDPi-ARD Workshop Cyberpolitics Workshop. Ahmed Shaheed and Benjamin Greenacre: Binary Threat: How Governments' Cyber Laws and Practice Undermine Human Rights in the MENA Region. *POMEPS Studies 43.* (2021).

[16] Afef Abrougui and Mohamad Najem. Follow the Money for Better Digital Rights in the Arab Region. *POMEPS Studies 43.* (2021).

[17] This distinction structures the export controls discussed in the following section.

[18] Navid Hassanpour, "Media Disruption and Revolutionary Unrest: Evidence From Mubarak's Quasi-Experiment," *Political Communication* 31, no. 1 (January 2, 2014): 1–24.

[19] See e.g. Netblocks, "Facebook Messenger, Social Media and News Sites Disrupted in Egypt amid Protests," *NetBlocks* (blog), September 22, 2019, https://perma.cc/K3WV-754U; Ronald Deibert, Joshua Oliver, and Adam Senft, "Censors Get Smart: Evidence from Psiphon in Iran," *Review of Policy Research* 36, no. 3 (2019): 341–56.

[20] Winnona DeSombre et al., "Countering Cyber Proliferation: Zeroing in on Access-as-a-Service" (Washington, D.C: Atlantic Council Cyber Statecraft Initiative, February 2021).

[21] P. N. Edwards, "Infrastructure and Modernity: Force, Time , and Social Organization in the History of Sociotechnical Systems," in *Modernity and Technology*, by Thomas J. Misa, Philip Brey, and Andrew Feenberg (Cambridge, Massachusetts: MIT Press, 2001), 185–226; Keller Easterling, *Extrastatecraft: The Power of Infrastructure Space*, Reprint edition (Verso Books, 2016).

[22] Nathan Ensmenger, "The Environmental History of Computing," *Technology and Culture* 59, no. 4 (2018): 514.

[23] See David Lyon, *The Culture of Surveillance: Watching as a Way of Life* (Cambridge, UK: Polity Press, 2018); Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power* (Profile Books, 2019).

[24] James Shires, *The Politics of Cybersecurity in the Middle East*, Chapters 3&4.
[25] Matteo Colombo, Federico Solfrini, and Arturo Varvelli, "Network Effects: Europe's Digital Sovereignty in the Mediterranean" (London, UK: European Council on Foreign Relations, May 4, 2021).
[26] Stephanie Kirchgaessner, "Revealed: Saudis Suspected of Phone Spying Campaign in US," *The Guardian*, March 29, 2020; for more detail on the vulnerability and exploitation, see Kim Zetter, "The Critical Hole at the Heart of Our Cell Phone Networks," *Wired*, April 28, 2016.
[27] Loqman Salamatian et al., "The Geopolitics behind the Routes Data Travels: A Case Study of Iran," *ArXiv:1911.07723 [Cs]*, November 19, 2019
[28] Olivier Tesquet, "Amesys: Egyptian Trials and Tribulations of a French Digital Arms Dealer," Telerama, July 5, 2017, https://perma.cc/L288-BMJN.
[29] But see online maps of Chinese investment in the Gulf released by the Australian Strategic Policy Institute (ASPI) for a more complex picture beyond two or three headline companies.
[30] Simeon Kerr, "UAE Caught between US and China as Powers Vie for Influence in Gulf," Financial Times, June 2, 2020; Narayanappa Janardhan, "Beijing Signals Growing Interest in Regional Conflict Management," *Arab Gulf States Institute in Washington*, December 22, 2020, https://perma.cc/YH2U-MRP6.
[31] Tesquet, "Amesys"; Al-Jazeera Investigative Unit. See also "Spy Merchants: Spying on Dissent through Illegal Means," *Al-Jazeera*, April 10, 2017, https://perma.cc/2CNY-EQR2.
[32] BBC, "How BAE Sold Cyber-Surveillance Tools to Arab States," *BBC News*, June 15, 2017, https://perma.cc/75ZM-NXYD; Winnona DeSombre et al., "Countering Cyber Proliferation: Zeroing in on Access-as-a-Service" (Washington, D.C: Atlantic Council Cyber Statecraft Initiative, February 2021).
[33] Staff Report, "Infinite Eyes in the Network: Government Escalates Attack on Secure Communication," Mada Masr, February 10, 2017, https://perma.cc/8DJ5-SHD9. Further attempts to block Signal in early 2017 inadvertently blocked all Google traffic to Egypt, causing outages and highlighting the technological limitations and experimental character of online censorship.
[34] Moxie Marlinspike, "A Letter from Amazon," Signal Messenger, May 1, 2018, https://signal.org/. See also Marlinspike's reflections in "Looking back at the front", on the same blog.
[35] Jakub Dalek et al., "Tender Confirmed, Rights at Risk: Verifying Netsweeper in Bahrain" (Citizen Lab, September 21, 2016); Nicki Thomas and and Amy Dempsey, "Guelph-Based Software Censors the Internet in the Middle East," *The Toronto Star*, June 13, 2011, https://perma.cc/KH8H-KXPT.
[36] Morgan Marquis-Boire et al., 'Planet Blue Coat: Mapping Global Surveillance and Censorship Tools' (Citizen Lab, January 2013).
[37] Bennett Haselton, 'Smartfilter: Miscategorization and Filtering in Saudi Arabia and UAE' (Citizen Lab, 28 November 2013).
[38] Privacy International, "The Global Surveillance Industry," July 2016; James Shires, *The Politics of Cybersecurity in the Middle East*, Chapter 4.
[39] Bill Marczak et al., "Bad Traffic: Sandvine's PacketLogic Devices Used to Deploy Government Spyware in Turkey and Redirect Egyptian Users to Affiliate Ads?" (Citizen Lab, March 9, 2018); Qurium Media Foundation, "How Operators Use Sandvine to Block Independent Media in Egypt," September 21, 2020, https://perma.cc/H6K5-5255; Ryan Gallagher, "Belarusian Officials Shut Down Internet With Technology Made by U.S. Firm," *Bloomberg.Com*, August 28, 2020.
[40] Christopher Bing and Joel Schectman, "Special Report: Inside the UAE's Secret Hacking Team of U.S....," *Reuters*, January 30, 2019, https://perma.cc/8DGA-95EQ
[41] Christopher Bing and Joel Schectman, "Mozilla Blocks UAE Bid to Become an Internet Security Guardian...," *Reuters*, July 9, 2019, https://perma.cc/P9J3-E42H.
[42] Moxie Marlinspike, "A Saudi Arabia Telecom's Surveillance Pitch," May 13, 2013, https://perma.cc/GT39-6JL5; Jenna McLaughlin, "Spies for Hire: How the UAE Is Recruiting Hackers to Create the Perfect Surveillance State," *The Intercept* (blog), October 24, 2016, https://perma.cc/5TXR-W8MV; Hagar Shezaf and Jonathan Jacobson, "Revealed: Israel's Cyber-Spy Industry Helps World Dictators Hunt Dissidents and Gays," *Haaretz*, October 20, 2018.
[43] James Shires, *The Politics of Cybersecurity in the Middle East,* Chapter 4.

The web (in)security of MENA civil society and media

Alexei Abrahams, University of Toronto[1]

Civil society and news media across the Middle East and North Africa (MENA), whether coordinating protests or subverting state-sanctioned discourse, increasingly rely upon digital communications to reach their audiences.[2] The majority of these communications, however, travel over infrastructure controlled by the same state and corporate authorities that civil society seeks to challenge. A growing body of scholarly evidence worldwide confirms that authorities exploit their 'man-in-the-middle' positions to interfere with civil society's communications.[3] Prominently, investigations by watchdog organizations like Citizen Lab or Amnesty International have proven that states hack the digital devices of activists and human rights defenders to surveil and repress them.[4] Within the Middle East, for example, there have been confirmed cases of digital surveillance of human rights defenders in the UAE, Saudi Arabia, Egypt, and Morocco.[5]

Are civil society and media in the MENA taking precautions to secure themselves against cyber attack? This question may seem natural to ask, but it has rarely been pursued by scholars. For a start, political scientists typically view security as a matter over which the state (the 'monopolist of violence') has chief prerogative, in which case the relevant question is not 'what is civil society doing to protect itself from attack?' but rather 'what is the state doing to protect civil society from attack?'[6] This question, of course, breaks down in situations where the state is itself the primary perpetrator of attacks against civil society. Indeed, in the MENA region, civil society is (correctly) viewed as a challenge to authoritarian control and is therefore routinely surveilled and repressed by the state, with digital surveillance and repression being merely the latest category of abuse. Shifting gears, one might turn to Western powers further up the international order to pressure the governments of Egypt, Saudi Arabia, or Israel, among others, to respect international law and human rights, or to restrict the export of cybersurveillance technologies to countries with a poor record in this regard. Indeed, this legal or rights-based approach is the starting premise of the work of organizations like Amnesty International, Human Rights Watch, or Access Now, and is implicit in several essays published in this collection.[7] Such an approach, however, must contend with the awkward fact that Western powers, far from living up to their professed values enshrined in human rights and international law, have a long-running history of arming, funding, and otherwise legitimizing abusive regimes across the MENA region.[8] In the absence of a clear moral and legal directive from Western governments, cybersecurity consultants to MENA regimes draw ethical boundaries according to their own professional or personal judgment.[9] But such idiosyncratic efforts do not and cannot amount to a hard guarantee of security for civil society from digital authoritarianism. Instead, civil society is left with no choice but to look to its own defenses. In this sense, then, we ask what civil society in the MENA is doing to protect itself from cyberattack.

So far, data and analysis have been scarce. Among cybersecurity professionals there is a profitable preoccupation with state and corporate security, and a concomitant neglect of civil society's security.[10] Cybersecurity itself, moreover, has tended to be conceptualized as *for* the state *from* decentralized actors -- not the other way around.[11] The past two decades, however, have witnessed a "moral maneuver" to recast cybersecurity according to a more human-centric paradigm, where civil society's security from state-level cyber intrusion is a tentatively legitimate object of inquiry.[12] Even so, researchers have thus far tended to favor threat reporting over security assessment, leaving us with some sense of attack capabilities but only piecemeal knowledge of defensive readiness. The handful of studies that do inquire about civil society's cybersecurity tend to ignore the MENA region, focus on individuals and ignore organizations, and ignore web security.[13] It is time that scholars of cyber politics in the Middle East begin to address these data gaps.

In this essay, I draw attention to an ongoing effort to collect and analyze data on the web security of civil society organizations (CSOs) and news media. While the literature has thus far dwelt primarily on the cybersecurity of individual activists and journalists, these individuals typically belong to CSOs, including charities, sustainable development NGOs, human rights NGOs, professional syndicates, labor unions, news agencies, and so on. Indeed, the coalescing of individuals around CSOs is an important part of the maturation process of social movements to convert ephemeral 'mobilization' into sustained 'organization'. CSOs increasingly maintain their own computer networks that both members and the broader community may regularly connect to and trust. Most saliently, CSOs (especially news agencies) increasingly run their own websites, hosted on web servers. These websites may often be the first and primary point of contact between a CSO and its community or audience. As social media platforms like Twitter and Facebook increasingly comply with local authorities to suppress civil society's speech,[14] independently managed websites are a natural fallback. All of this highlights the importance of knowing whether CSO web infrastructure is secure against cyber attack.

Web scanning

For researchers, an advantage to studying CSO web security is that websites are (by design) easy to find and can be scanned remotely. By contrast, individuals' devices are harder to enumerate, may require greater ethical precautions (possibly including obtaining consent from the individual), and may necessitate in-person assessments (see Marczak and Paxson 2017). In ongoing work, I and a coauthor script a custom web scanning tool to remotely gather web security information from CSO websites.[15] The tool, while by no means constituting a formal security assessment, yields a variety of data highly relevant to measuring a website's security posture and, by extension, the security of the organization.

For a start, we check whether CSOs offer or insist on encrypted web sessions via the HTTPS protocol. Failure to encrypt implies that a visitor's session can be read in plaintext, and tampered with, by state authorities with man-in-the-middle (MITM) positions at any intermediate 'hop' between the visitor's device and the web server.

Websites can also be overwhelmed with malicious traffic in distributed denial of service (DDoS) attacks. We scan sites to see if they have implemented caching mechanisms to frustrate such attacks.

Nowadays most websites are typically not built from scratch, but rather draw upon third-party templates or frameworks such as WordPress, Drupal, and Joomla. As security vulnerabilities are detected, content management system (CMS) companies issue updates, but a CSO may fail to stay abreast of these updates. We fingerprint the website's CMS and check if it is up-to-date.

Even if the website software is up-to-date, the underlying web server's software may not be. We augment our website scan with readouts from Shodan to see if web servers are running software with publicly known vulnerabilities.

Finally, we identify the country where the web server is geolocated. If the CSO's state adversary has physical access to the server, then its security may be compromised.

Results

Palestine

As a proof of concept, Abrahams and Anonymous (2021) scan Palestinian civil society, non-Palestinian CSOs openly allied with the Boycott, Divestment, and Sanctions (BDS) movement, and a benchmark sample of Israeli think tanks and news agencies. In view of its contentious political history with Israel, Israel's status as a world-renowned cyber threat actor, and Israel's 'man-in-the-middle' position on internet traffic in the West Bank, Palestinian civil society arguably constitutes something of a paradigmatic case for thinking about the cybersecurity of civil society. The results of their scan are reprinted below in Table 1.

Table 1: Comparing security of Palestinian, BDS, and Israeli websites

	Civil society websites (news agencies, think tanks, CSOs)		
	Palestinian	**BDS**	**Israeli**
Allow https sessions	**68.0%** (155/228)	**74.0%** (54/73)	**74.4 %** (64/86)
Force https sessions	**41.7%** (95/228)	**54.8%** (40/73)	**54.7%** (47/86)
Up-to-date CMS*	**33.7%** (28/83)	**27.5%** (11/40)	**52.2%** (12/23)
X-Frame-Options	**13.2%** (30/228)	**5.5%** (4/73)	**8.1%** (7/86)
Strict-Transport-Security	**5.7%** (13/228)	**19.2%** (14/73)	**5.8%** (5/86)
Content-Security-Policy	**1.3%** (3/228)	**1.4%** (1/73)	**7.0%** (6/86)
DDoS protection**	**0.4%** (1/228)	**9.6%** (7/73)	**17.4%** (15/86)
No high/critical CVEs***	**77.2%** (176/228)	**78.1%** (57/73)	**81.2%** (70/86)

**For technical reasons, the CMS version for each website could not always be identified.*

***Detection was limited to the use of Cloudflare, Google Cloud or Deflect for DDos protection, three solutions offered for free to civil society but that do not together constitute an exhaustive list of DDoS mitigations.*

****This number is calculated by querying the Shodan API (https://shodan.io).*

Insecurity of Palestinian CSO web infrastructure

The first column of Table 1 summarizes cybersecurity statistics for a sample of 228 Palestinian CSOs. In absolute terms, the security of these organizations leaves a lot to be desired. Roughly a third of Palestinian CSO websites disallow encrypted sessions, only four in ten insist on them, and less than 6% enforce strict transport security. These lapses invite state authorities to monitor and modify visitor sessions, possibly even delivering malicious payloads. Fully two thirds of CSO websites are running outdated versions of WordPress (45), Joomla (5), and Drupal (5). Likewise, the underlying servers of 52 (22.8%) websites have at least one vulnerability rated 'high' or 'critical'. Using outdated software is dangerous since software updates often patch publicly disclosed security vulnerabilities that can otherwise be exploited by attackers.

What explains these security lapses? The answer can neither be straightforwardly technical nor financial. For example, many Palestinian CSOs use plaintext HTTP, but the secure protocol HTTPS has existed for over two decades, and upgrading is a free and fairly seamless process.[16] Similarly, DDoS protection through Cloudflare, Google Cloud, and Deflect, is offered free-of-charge for civil society organizations. By definition, outdated web server and website software can be updated, typically for free. Nor can one claim that Palestinians organizations are unaware of these solutions; indeed, many Palestinian organizations have implemented them even as many have not.

Could it be that these security lapses are idiosyncratic to life under military occupation, or some other idiosyncratic challenge faced by Palestinian civil society? To shed light on this, Abrahams and Anonymous (2021) scanned 86

Israeli think tanks and news agencies. Column 2 of Table 1 lists the results. While Israeli CSO websites do appear to perform better on most security metrics, in absolute terms they are quite insecure, too. To take just one example, 74.4% of Israeli websites allow encrypted web traffic, which is only marginally higher than Palestinian CSOs (68.0%).

Finally, one might argue that Palestinian CSOs neglect their web security because they enjoy a 'security by obscurity' insofar as they remain disengaged from politically contentious action.[17] Indeed, scholars have argued that Palestinian civil society has become de-politicized and co-opted to a neoliberal development agenda since the start of the Oslo period, implying that they are an unlikely target of Israeli cyber surveillance.[18]

To evaluate this possibility, Abrahams and Anonymous (2021) looked beyond their sample of Palestinian CSOs to consider organizations belonging to the Boycott, Divestment, and Sanctions (BDS) movement. The BDS movement is undoubtedly contentious, and has drawn all manner of hostility, including cyberattacks in recent years.[19] They scan 73 non-Palestinian organizations publicly affiliated with the BDS movement. The results, listed in Column 3, suggest their web infrastructure appears to be no more prepared for attack than 'ordinary' Palestinian CSOs. Indeed, over five years since the BDS movement's main website suffered a DDoS attack,20 they find that just 9.6% of BDS-affiliated organizations have availed themselves of DDoS protection. Only a quarter of BDS-affiliated organizations run websites with up-to-date software. And the websites for one in five organizations are hosted on web servers running out-of-date software for which high/critical vulnerabilities are publicly known. The decision to openly engage in contentious political action, it would seem, does not prompt these organizations to adopt a higher degree of security vigilance.

Media websites across the MENA region

One potential explanation for the insecurity of CSO websites is that they may have been launched perfunctorily to please donors and are largely irrelevant to the day-to-day operations of these organizations. While this may hold true for many CSOs, the argument does not easily extend to news media. Media websites are often the first and primary point of contact between a news agency and its audiences. And since journalists are themselves often the targets of surveillance, it stands to reason that news organizations – as central points of contact within media networks – would likewise be targeted.

I therefore widened the aperture of the scan to news media sites across the region. I obtained a list of news websites from MediaCloud's geographic collections for each of 20 MENA countries, then pointed the web scanning tool at them. At this preliminary stage, I am not yet able to disaggregate between state-aligned versus independent news organizations, nor do I yet give greater weight to more popular sites. With those caveats in mind, Table 2 lists the web security statistics averaged across the twenty MENA countries scanned, while Figures 1 and 2 depict cross-country comparisons for two metrics (HTTPS availability and web server vulnerabilities).

Table 2: Web security of media sites across the MENA

	Average across all MENA countries
Allow https sessions	80.0%
Force https sessions	58.5%
Up-to-date CMS*	50.6%
X-Frame-Options	10.2%
Strict-Transport-Security	8.2%
Content-Security-Policy	3.5%
DDoS protection**	34.4%
No high/critical CVEs***	86.2%

**For technical reasons, the CMS version for each website could not always be identified.*

***We only detected the use of Cloudflare, Google Cloud or Deflect for DDos protection, three solutions offered for free to civil society but that do not together constitute an exhaustive list of DDoS mitigations.*

****This number is calculated by querying the Shodan API (https://shodan.io).*

Figure 1: Percentage of media websites per country offering HTTPS

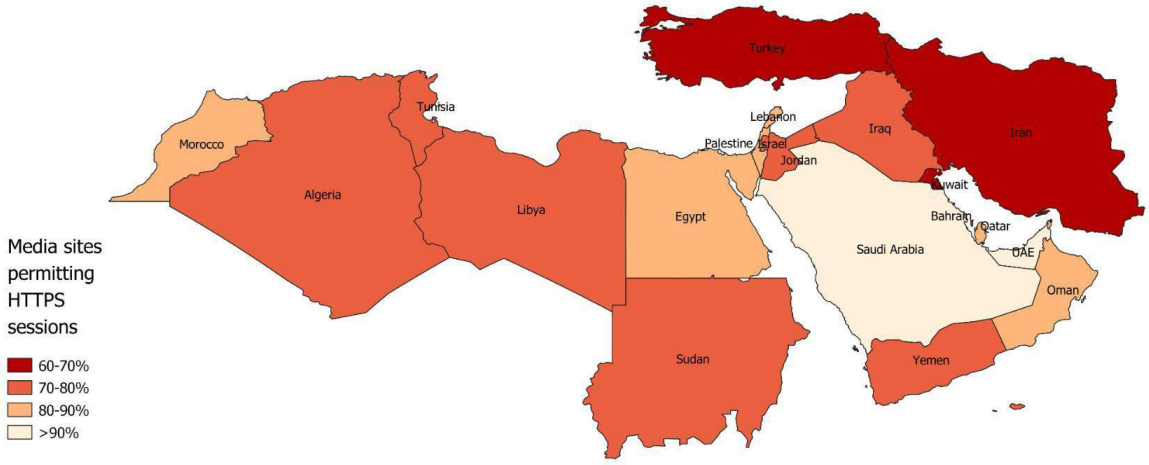
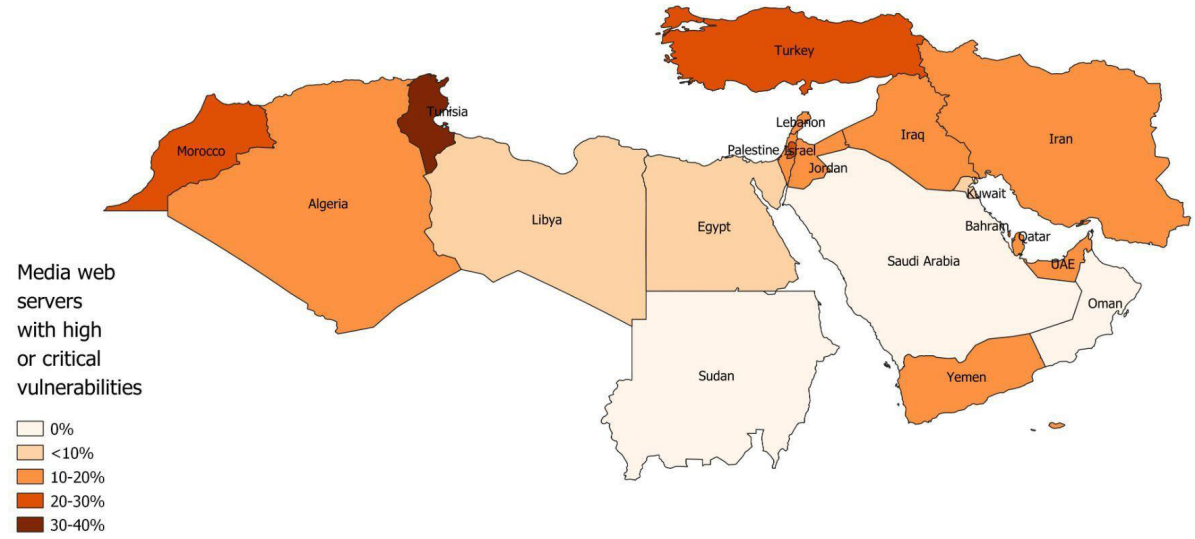


Figure 2: Percentage of media web servers per country running software with a 'high' or 'critical' vulnerability (CVSS of 7.0 or above)



Overall, despite the importance of websites for news media organizations, the results in Table 2 suggest that while they do exhibit better web security than the CSO organizations, they still leave a lot to be desired. Averaged across countries, 80% of media sites in the MENA enable HTTPS. While Kuwaiti media sites score the worst on this metric (66.7%), Saudi Arabia's media sites exhibit the highest rates of encryption, (92.9%) and also force encryption at the highest rate regionally (85.7%). Saudi Arabian media also lead the region with the lowest rates of web server vulnerabilities -- *none* of their media web servers exhibits high or critical vulnerabilities. On the other end of the spectrum, 31.6% of Tunisian media sites have high or critical web server vulnerabilities. Across the region, one in three media sites enjoys DDoS protection, with Egyptian media the best protected (74.3%).

The cross-country variation in these data invites further investigation. Saudi Arabia's high scores may reflect its investments in internet technologies more generally, while Egypt's status as a digital hub within the region may have something to do with its higher rates of DDoS protection.[21] As the region's only nascent democracy, Tunisia's vulnerable media web servers are troubling, and suggest there may not be a straightforward relationship between governance type and media security. Further investigation will be required to unpack each of these differences, but these data constitute a jumping-off point.

Conclusion

In the wake of the Arab Spring, citizens across the MENA region continue to demand accountable governance and challenge state and corporate authority. Within this conflict, civil society and media are a crucial mesolayer between citizens and the state, helping citizens to stay informed and collectivized. These vital roles of civil society and media make them a target of co-option and repression by the state. As digital communications have gained importance for civil society and media, states have stepped up cyber surveillance and interference. Despairing of intercession from Western powers, civil society and media in the region must look to their own defenses.

How are civil society and media prepared to meet these cyber threats? In this essay, I drew attention to an ongoing effort to scan the web infrastructure of civil society and media organizations. Worrisomely, the scans reveal widespread and potentially compromising insecurities. On the other hand, in all cases technical solutions exist and can generally be implemented easily and at minimal cost. Positively, then, and in contrast to the gloomy forecasts surrounding the rise of digital authoritarianism in the region, it would appear that there is much that civil society and media in the region can do unilaterally to protect themselves.

At the same time, scholars of cyber politics can do more to fill the data and analytical lacuna around this topic. The web scan results surface some puzzling similarities and differences between countries and sectors and invite further research. Moreover, the persistence of these insecurities defies straightforward technical, financial, or contextual explanations, and ought to prompt deeper investigation, likely including interviews of the organizations themselves. For comparative perspective, the aperture of these scans should be widened to include different sectors of society (civil society, media, private sector, government) to clarify the full range and depth of the situation. On the other hand, web-facing infrastructure is only one 'attack surface' among many. Further efforts along the lines of Marczak and Paxson (2017) to assess the security of individuals' devices and unpack individuals' decision making would be invaluable. The cybersecurity of civil society and media is a topic of emerging importance to the region, and scholars have both an opportunity and a responsibility to get involved.

Endnotes

[1] Postdoctoral fellow, Technology & Social Change Project, Shorenstein Center, Harvard University, alexei_abrahams@hks.harvard.edu

[2] Larry Diamond, & Marc F. Plattner (Eds.), *Liberation technology: Social media and the struggle for democracy* (Baltimore: Johns Hopkins University Press, 2012). Philip N. Howard and Muzammil M. Hussain, *Democracy's fourth wave?: digital media and the Arab Spring* (Oxford: Oxford University Press, 2013). Manuel Castells, *Networks of outrage and hope: Social movements in the Internet age* (Cambridge: Polity, 2012).

[3] Marc O. Jones, *Digital Authoritarianism in the Middle East* (forthcoming, London: Hurst, 2021). Margaret E. Roberts, *Censored: distraction and diversion inside China's Great Firewall* (Princeton, NJ: Princeton University Press, 2018). Nils B. Weidmann and Espen G. Rød, *The Internet and political protest in autocracies* (Oxford, UK: Oxford University Press, 2019).

[4] Amnesty International, "Click and Bait: Vietnamese Human Rights Defenders Targeted with Spyware Attacks" (2021), available from: https://www.amnesty.org/en/latest/research/2021/02/click-and-bait-vietnamese-human-rights-defenders-targeted-with-spyware-attacks. John Scott-Railton, Bill Marczak, Siena Anstis, Bahr AbduRazzak, Masashi Crete-Nishihata, and Ron Deibert, "Reckless VII: Wife of Journalist Slain in Cartel-Linked Killing Targeted with NSO Group's Spyware" (2019), available from: https://citizenlab.ca/2019/03/nso-spyware-slain-journalists-wife

[5] Bill Marczak and John Scott-Railton, "The million dollar dissident: NSO group's iPhone zero-days used against a UAE human rights defender", The Citizen Lab (2016). Bill Marczak, John Scott-Railton, Adam Senft, Bahr AbduRazzak, and Ron Deibert, "The Kingdom Came to Canada - How Saudi-Linked Digital Espionage Reached Canadian Soil" (2018), available from: https://citizenlab.ca/2018/10/the-kingdom-came-to-canada-how-saudi-linked-digital-espionage-reached-canadian-soil. John Scott-Railton, Bill Marczak, Ramy Raoff, and Etienne Maynier, "Nile Phish - Large-Scale Phishing Campaign Targeting Egyptian Civil Society" (2017), available from: https://citizenlab.ca/2017/02/nilephish-report. Amnesty International, "Morroccan Journalist Targeted with Network Injection Attacks using NSO Group's Tools" (2020), available from: https://www.amnesty.org/en/latest/research/2020/06/moroccan-journalist-targeted-with-network-injection-attacks-using-nso-groups-tools, Amnesty International, "German-made FinSpy spyware found in Egypt, and Mac and Linux versions revealed", (2020), available from: https://www.amnesty.org/en/latest/research/2020/09/german-made-finspy-spyware-found-in-egypt-and-mac-and-linux-versions-revealed.

[6] Alexei Abrahams and Brandon Merrell, "Monopolies of Violence? How Insurgent Threats Can Motivate Accountable Governance," 2021.

[7] Ahmed Shaheed and Benjamin Greenacre, "Binary Threat: How Governments' Cyber Laws and Practice Undermine Human Rights in the MENA Region", POMEPS Studies (2021). Mohamad Najem and Afef Abroughi, "Follow the Money for Better Digital Rights in the Arab region", POMEPS Studies (2021).

[8] Among myriad books and articles, see for example Amaney Jamal, *Of empires and citizens: pro-American democracy or no democracy at all?*. Princeton University Press, 2012. Jason Brownlee, *Democracy prevention: The politics of the US-Egyptian alliance*. Cambridge University Press, 2012. Sean L.Yom, *From resilience to revolution: How foreign interventions destabilize the Middle East*. Columbia University Press, 2015.

[9] James Shires, "The implementation of digital surveillance infrastructures in the Gulf", POMEPS Studies (2021).

[10] Lennart Maschmeyer, Ron Deibert, and Jon Lindsay, "A tale of two cybers - how threat reporting by cybersecurity firms systematically underrepresents threats to civil society", *Journal of Information Technology & Politics,* Vol 18 (2020).

[11] James Shires, *The politics of cybersecurity in the Middle East* (forthcoming, London: Hurst, 2021).

[12] Ibid.

[13] Two studies of civil society cybersecurity in the MENA region are Bora Ataman and Barış Çoban, "Counter-surveillance and alternative new media in Turkey", *Information, Communication & Society* 21, no. 7 (2018): 1014-1029; and Bill Marczak and Vern Paxson, "Social Engineering Attacks on Government Opponents: Target Perspectives", *Proceedings on Privacy Enhancing Technologies*, *2* (2017), 172-185. One study that focuses on organizations is Nikita Samarin, Alisa Frik, Sean Brooks, Coye Cheshire, and Serge Egelman, "Conducting Privacy-Sensitive Surveys: A Case Study of Civil Society Organizations", *arXiv preprint arXiv:2003.08580* (2020).

[14] Mona El Swah and Mahsa Alimardani. "Digital Apartheid: #SaveSheikhJarrah and Arabic Content Moderation". POMEPS Studies, 2021. Systematic Efforts to Silence Palestinian Content On Social Media. https://7amleh.org/2020/06/07/systematic-efforts-to-silence-palestinian-content-on-social-media

[15] Alexei Abrahams and Anonymous, "Measuring the (in)security of Palestinian civil society web infrastructure", 2021.

[16] https://letsencrypt.org/

[17] Susan E. McGregor, Elizabeth A. Watkins, "'Security by Obscurity': Journalists' Mental Models of Information Security", *ISOJ* Vol 6:1 (2016).

[18] Mona Atia and Catherine E. Herrold, "Governing through patronage: The rise of NGOs and the fall of civil society in Palestine and Morocco", *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *29*(5), 1044-1054 (2018). Tariq Dana, "The structural transformation of Palestinian civil society: Key paradigm shifts", *Middle East Critique*, *24*(2), 191-210 (2015). Benoit Challand, *Palestinian civil society: Foreign donors and the power to promote and exclude*, (New York, NY: Routledge, 2008). Amaney Jamal, *Barriers to democracy: The other side of social capital in Palestine and the Arab world*, (Princeton, NJ: Princeton University Press, 2009).

[19] Nathan Brown and Daniel Nerenberg, "Palestine in Flux: From Search for State to Search for Tactics," *Carnegie Endowment for International Peace* (2016). Available from https://carnegieendowment.org/2016/01/19/palestine-in-flux-from-search-for-state-to-search-for-tactics-pub-62486. eQualitie, "Deflect Labs Report #2 : Botnet Attack Analysis of Deflect Protected Website bdsmovement.net", (2016) available from: https://equalit.ie/en/deflect-labs-report-2

[20] eQualitie (2016)

[21] See Shires (2021) and Jones (2021) for background.

Beyond Liberation Technology?

The Recent Uses of Social Media by Pro-Democracy Activists[1*]

Joshua A. Tucker, New York University [2**]

It is difficult to believe it, but it has been almost a decade since the publication of Larry Diamond and Marc Plattner's *Liberation Technology*.[3] In that period of time, we have gone from viewing social media as an exciting new phenomenon with the potential to fundamentally change the nature of political protest around the globe – a phenomenon which had to be painstakingly identified, labeled, and assessed in each new protest movement – to a world where it is practically impossible to imagine a protest movement that doesn't involve some, if not many, aspects of social media in its planning, implementation, and coverage, to say nothing of the regime's response to that protest. At the same time, the speed and pace of change in the digital information environment seems to have accelerated to such a degree that we now have calls from social scientists for the importance of assessing the *temporal validity* of research studies.[4] One would suppose, therefore, that the nature of the relationship between social media and pro-democracy activism might have fundamentally changed over that period.

Counter-intuitively, then, perhaps, I argue the opposite. For the most part, the theoretical frameworks we have developed over the past decade to explain the relationship between social media and pro-democracy movements actually do a relatively good job in giving us the tools to understand what has been happening in recent pro-democracy movements outside the Middle East in places such as Belarus, Russia, Hong Kong, and Myanmar. Yes, there are new platforms (Telegram, YouTube, TikTok) that have risen to places of prominence in pro-democracy movements. And yes, there are new forms of media (video, group chats) that are now utilized in ways they were not previously, both by pro-democracy movements and the regimes against which they mobilized. But the purposes to which these tools are put, and the functions which they seem to be fulfilling, are remarkably similar to the purposes and functions ascribed to social media in pro-democracy movements in the first half of the previous decade.

To make these points, I organize the remainder of the essay as follows. First, I begin with a summary of a general theoretical framework for thinking about social media and protest drawn from three of my previous publications with different sets of co-authors on social media's relationship to protest and pro-democracy movements in the first half of the last decade.[5] Next, I turn to the ways in which pro-democracy activists have used new platforms that were not as prevalent in our original Twitter and Facebook narratives of past protests, but I will argue that they are being used for essentially similar functions. I will then briefly address the ways that regimes have responded to recent protest movements, again arguing that existing theoretical frameworks can account for these types of responses. I will close, however, with what I think is a genuinely new development that is not accommodated by our prior theoretical frameworks: the fact that the platforms themselves have in some cases taken sides in these conflicts. In all cases, I draw upon reports from other scholars and journalists regarding recent development over the past year or two in Belarus, Russia, Myanmar, and Hong Kong.[6]

Theoretical Frameworks for Social Media and Pro-Democracy Activists

I will begin by positing that our current understanding of the uses of social media by pro-democracy activists can be (somewhat) concisely summarized by the following the propositions:

First, social media can serve to circumvent authoritarian rule by giving voice to those without access to mainstream

media.[7] In countries ruled by authoritarian – or competitive authoritarian[8] – regimes, those without access to mainstream media may include other would-be authoritarians, but also undoubtedly includes pro-democracy forces. Thus, social media provides tools to pro-democracy activists that would not likely otherwise be available in the absence of social media.

Second, social media can be used in a variety of formats by pro-democracy activists that are directly connected to mass protest. This includes social media being used to *organize* protest events, to communicate *real-time information* about protests to protests participants, to *drive media coverage* of protests events, as well as to *build networks* of pro-democracy activists.[9]

Third, although social media clearly provides tools to pro-democracy activists, regimes are not powerless to respond to opposition on social media. Indeed, regimes possess a variety of options for doing so, including offline responses (such as arresting online activists or changing the liability or ownership structure of platforms), online restriction of access to information (think traditional attempts at censorship), and attempts to influence the nature of the online discussion (including actors such bots and trolls).[10]

As these arguments were laid out across articles published in 2017 and 2018, it is a legitimate question to ask whether or not they continue to adequately account for social media's usage in pro-democracy struggle in recent years? To answer this question, I turn next to what I would argue are the two most important developments in the use of social media by pro-democracy activists outside of the Middle East in recent years: the rise of encrypted chat apps, and in particular Telegram; and the growing prominence of platforms that privilege images and video as opposed to text.

Telegram

Telegram, founded by VKontakte founder Pavel Durov with his brother Nikolai, is a kind of cross between WhatsApp and Twitter, in so far as it has the ability for one-on-one or group encrypted chats (as well as voice and video calls), but also provides for one-to-many communication through "channels" that users can follow in a way similar to following another account on Twitter. Telegram is also known for its privacy functions, including an even more secure "secret chat" function.[11]

The rise of the use of Telegram has been perhaps the biggest change in the use of social media by pro-democracy activists in the past couple of years. In Hong Kong, for example, Telegram has been credited (along with Signal and WhatsApp) with facilitating the protesters' "Flash Mob" strategy through enabling coordinated real time communication.[12] Alexandra Urman and co-authors argue that Telegram's privacy features made it particularly attractive for Hong Kong protesters aiming to avoid detection and retribution; they also present evidence from content analysis of public telegram channels that the platform was used to distribute information about protest times, locations, and police presence.[13] One creative use of Telegram in Hong Kong was to use the platform to allow protest participants to vote in real time about protest strategy.[14]

Telegram has also proved quite popular among participants in the past year's pro-democracy movement in Belarus, with it being described as "basically the only means of connecting to the internet" during attempts by the Belarusian state to shut down access to the internet.[15] One interesting feature of Telegram that has seemed consequential for recent pro-democracy protesters has been the ability to set up chat groups based on location. According to political scientists Aliaksandr Herasimenka's analysis of Telegram activity in the 2020 Belarus protests, most of the protest leaders were anonymous individuals who created these local Telegram groups; indeed, Herasimenka's assessment was that there was actually fairly little coordination of protest activity beyond these local groups.[16] Fascinatingly, someone set up an interactive map (dze.chat) that shows different local Belarusian Telegram groups by location that users can join:

Figure 1. Belarusian Localized Telegram Groups: Screen shots from Dze.chat on 5/13/21.

While Telegram is clearly "new" in the sense that it did not even exist at the time of Arab Spring and it is apparently taking on greater prominence in recent events, the role it is playing is exceedingly similar to that played by older social media platforms in prior pro-democracy movements: providing pro-democracy activists with a way to communicate and recruit participants without having to work through mainstream media and facilitating real-time communication during protest events. Of course, while the ability to communicate privately (and anonymously) with small groups of people that can easily organize based on location may be somewhat different than the earlier days of social media inspired protest, it is striking how the function played by these forms of communication remains very much in line with the role played by Twitter in Gezi Park or Facebook in Tahir Square. Further, as much as Telegram has become a useful tool of pro-democracy movements, it is worth being aware of the fact that it has reportedly become similarly popular among far-right movements in the United States.[17] This nicely illustrates the point we made in our *Journal of Democracy* article "From Liberation to Turmoil: Social Media and Democracy": the very same affordances of social media that make it attractive for pro-democracy activists can also make it attractive for anti-democratic forces in open societies.[18]

Video and Images

The second new development in recent pro-democracy movements has been the rise of the use of images and video, and in particular social media platforms that have affordances that feature images and video. So for example, in February, Russians posted images of themselves on Instagram wearing red clothing as a sign of support for opposition leader Alesei Navalny and his wife Yulia, as she had worn red during her husband's trial.[19] Russia was also the location of a series of protests on TikTok related to Navalny, where, according to the *Moscow Times*, videos with the hashtags "Free Navalny" and "Jan.23" had over 50 million combined views.[20] The content of the videos were clever and in keeping with the ethos of TikTok being fun to watch: one set of videos showed young packing to go attend protests; another set involved people

removing Putin's portrait from the walls of their schools and replacing them with pictures of Navalny.[21] Navalny's return to Russia in January of 2021 was accompanied by the release of his latest corruption exposé on YouTube, the "Putin's Palace" video that had been viewed over 115 million times at the writing of this essay.[22]

While featuring the use of video and image friendly platforms that were either not present or less popular in the original waves of pro-democracy movements on social media, social media is still playing a similar role in these movements. As Russian opposition politician Vladimir Kara-Murza noted, "this modern technology gives us the tools to counteract and stand up to the massive machine for examples of government propaganda."[23] The Navalny protest videos played a familiar role in supporting the protest movement by allowing would-be protesters to know that they would not be protesting alone, harkening back to arguments made decades ago by Timur Kuran[24] about the role played by knowing how many other people support a pro-democracy movement before deciding whether or not to join yourself.[25] Additionally, Navalny's use of YouTube to broadcast what in other times and places would have been a TV special shows how social media provides a broadcast platform to those denied access to traditional media.

Regimes Respond and the Cat and Mouse Games Continue

Just as was the case in earlier rounds of pro-democracy movements, regimes have responded to protesters' use of social media to bolster their cause. Recent developments in Myanmar provide good examples. As pro-democracy protesters attempted to use social media such as Facebook to "share information to create international awareness" of the situation on the ground,[26] the regime responded by using both heavy handed measures to restrict access to social media by essentially shutting down the internet at regular intervals for extended periods of time in February, 2021, as well as, supposedly, more surgically precise measures such as "blocking individual SIM cards that were believed to be in use by activists".[27] Moreover, in Myanmar members of the military actually took to TikTok to issue death threats to protesters, making their own videos where they showed themselves brandishing large guns.[28] While soldiers making TikTok videos is new, the fact that the regime would respond to online opposition by using social media to try to silence opposition voice is exactly what political scientist Margaret Roberts referred to as the regime's "Fear" tactic.[29]

What does appear to be genuinely new and outside of the existing framework of thinking about social media platforms as neutral arenas in which political actors compete with one another,[30] however, is that in Myanmar the platforms themselves do appear – however tentatively – to be taking sides. So, the TikTok video of soldiers issuing death threats? TikTok issued an announcement that it would remove the videos of soldiers issuing death threats.[31] Even more dramatically, Facebook banned the Myanmar military from the platform on March 3, 2021, a little more than two months after Donald Trump had been indefinitely kicked off the platform.[32] As Marwa Fatfata, Ahmad Shaheed and others show in this collection, Facebook and Instagram removed significant amount of pro-Palestinian content during the most recent crisis with Israel.

Of all the recent developments in the use of social media in pro-democracy movements, this one may be the one worth watching the most in the future. While it is clear that pro-democracy movements will find creative ways to utilize new platforms and new affordances of these platforms, for now it seems like the goals to which they are putting social media – publishing content that mainstream media would not have published, planning protest actions and sharing information about them in real time, and driving international media coverage – in recent arenas such as Belarus, Hong Kong, Russia, and Myanmar seems fairly similar to those from the first wave of social media assisted protests. But if the platforms themselves are going to be increasingly weighing in support of – or in opposition to -- these movements or their targets, then that might be a consequential change worth watching closely in the future.

Endnotes

[1] * Essay prepared for presentation at the "Digital Activism and Authoritarian Adaptation in the Middle East" conference jointly hosted by Stanford University CDDRL and POMEPS. I thank Trellace Lawrimore for helpful research assistance, and am grateful for the many excellent co-authors I had on the previous papers summarized in the first section of this essay: Megan Metzger, Sergey Sanovich, Denis Stukal, Pablo Barbera, Molly Roberts, and Yannis Theocaris.

[2] ** Professor of Politics, Affiliated Professor of Russian and Slavic Studies, Affiliated Professor of Data Science, Director, Jordan Center for the Advanced Study of Russia, and Co-Director, Center for Social Media and Politics (csmapnyu.org), New York University. Email: joshua dot tucker at nyu dot edu; Twitter: @j_a_tucker.

[3] Larry Diamond and Marc F. Plattner, eds. *Liberation technology: Social media and the struggle for democracy*. JHU Press, 2012.

[4] Kevin Munger. "Knowledge decays: Temporal validity and social science in a changing world." *Unpublished manuscript. https://osf. io/4utsk* (2019).

[5] Megan MacDuffee Metzger and Joshua A. Tucker. "Social media and EuroMaidan: A review essay." *Slavic Review* 76, no. 1 (2017): 169-191. Tucker, Joshua A., Yannis Theocharis, Margaret E. Roberts, and Pablo Barberá. "From liberation to turmoil: Social media and democracy." *Journal of democracy* 28, no. 4 (2017): 46-59. Sanovich, Sergey, Denis Stukal, and Joshua A. Tucker. "Turning the virtual tables: Government strategies for addressing online opposition with an application to Russia." *Comparative Politics* 50, no. 3 (2018): 435-482.

[6] We have not conducted original research on any of these conflicts at the NYU Center for Social Media and Politics, so the examples given in this essay – at the direction of the conference organizers – all represent secondary assessment of reports from other scholars and journalists.

[7] Joshua A. Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. "Social media, political polarization, and political disinformation: A review of the scientific literature." *Political polarization, and political disinformation: a review of the scientific literature (*March 19, 2018*)* (2018).

[8] Steven Levitsky and Lucan A. Way. *Competitive authoritarianism: Hybrid regimes after the Cold War*. Cambridge University Press, 2010.

[9] See Metzger and Tucker 2017, 179-187. The last of these functions – building networks – is probably the most contested. The very fact that social media usage can build online networks is not in debate, but arguments persist about the durability of these networks compared to traditional in-person networks. For example, the TikTok protest video in Russia described later in this essay were assessed to be relatively short-lived with little legacy effect in the form of establishing a linked network by political scientists Alexandra Urman this is a probably a good example of the importance of paying attention to the *affordances* of social media platforms. (*Navalny and the Kremlin: Politics and Protest in Russia*. New York City-Russia Public Policy Series. NYU Jordan Center for the Advanced Study of Russia and Harriman Institute at Columbia University, February 1, 2021. https://www.youtube.com/ watch?v=BUfcxe8BT68). Facebook, for example, is very well suited for growing networks due to the "Groups" and "Pages" features in a way that Twitter, TikTok, and even What's App (due to maximum groups sizes) may not be. In contrast, Telegram seems well positioned to do so due to its ability to facilitate communication among large private groups.

[10] Sanovich et al. 2018. For a closely related alternative classification, see Roberts, Margaret *Censored*. Princeton University Press, 2018, where Roberts categories the options of regimes as the "three Fs": fear, friction, and flooding.

[11] For more details, see https://en.wikipedia.org/wiki/Telegram_(software) or https://www.androidauthority.com/what-is-telegram-messenger-979357/.

[12] Tin-yuet Ting. "From 'be water'to 'be fire': nascent smart mob and networked protests in Hong Kong." *Social Movement Studies* 19, no. 3 (2020): 362-368. Flash mob protests have also been used in Belarus: https://www.rferl.org/a/belarusian-protests-continue-using-flash-mob-tactics-to-avoid-police-crackdown/31039954.html.

[13] Aleksandra Urman, Justin Chun-ting Ho, and Stefan Katz. ""No Central Stage": Telegram-based activity during the 2019 protests in Hong Kong." (2020) https://osf.io/preprints/socarxiv/ueds4/.

[14] https://www.bbc.com/news/technology-48802125

[15] https://cointelegraph.com/news/telegram-supports-protests-in-belarus-with-new-opposition-flag-emoji.

[16] *Belarus: Looking Forward and Looking East to Russia*. New York City-Russia Public Policy Series. NYU Jordan Center for the Advanced Study of Russia and Harriman Institute at Columbia University, December 8, 2020. https://www.youtube.com/watch?v=Ka4x1h8qZfs.

[17] https://www.propublica.org/article/this-is-war-inside-the-secret-chat-where-far-right-extremists-devised-their-post-capitol-plans.

[18] Tucker et al. (2017)

[19] https://www.nytimes.com/2021/02/09/style/aleksei-navalny-red-supporters.html

[20] https://www.themoscowtimes.com/2021/01/21/young-russians-flood-tiktok-with-navalny-protest-cries-a72676

[21] https://www.themoscowtimes.com/2021/01/21/young-russians-flood-tiktok-with-navalny-protest-cries-a72676

[22] https://www.youtube.com/watch?v=ipAnwilMncI&t=3441s

[23] Nathan Hodge, Vladimir Kara-Murza, and Andrei Soldatov. Inside Russia's Protest Movement. Online broadcast. Chicago Council on Global Affairs, March 31, 2021. https://www.youtube.com/watch?v=kJJg53b7cD4.

[24] Timur Kuran. "Now out of never: The element of surprise in the East European revolution of 1989." *World politics* 44, no. 1 (1991): 7-48.

[25] Point made by Alexandara Urman during the Q&A of *Navalny and the Kremlin: Politics and Protest in Russia*. New York-Russia Public Policy Series. NYU Jordan Center for the Advanced Study of Russia and Harriman Institute at Columbia University, February 1, 2021. https://www.youtube.com/watch?v=BUfcxe8BT68.

[26] How the internet has redefined protest in Myanmar. Online. CNN, March 24, 2021. https://www.cnn.com/videos/world/2021/03/24/myanmar-protests-internet-facebook-lon-orig.cnn.

[27] Doug Bock Clark. "Internet Access Complicates the Coup in Myanmar". The New Yorker (March 2, 2021). https://www.newyorker.com/tech/annals-of-technology/in-myanmar-a-digital-savvy-nation-poses-a-new-challenge-for-the-military. For more on internet shutdowns globally, see "Woodhams, Samuel and Simon Migliano. The Global Cost of Internet Shutdowns in 2020. Top10VPN, January 4, 2021. https://www.top10vpn.com/cost-of-internet-shutdowns/SamuelWoodhams&SimonMigliano and Jamjoom, Mohammed. Will internet shutdowns become the norm? — Inside Story. Online broadcast. Al Jazeera English, February 14, 2021. https://www.youtube.com/watch?v=a3V_dey7Fdk&list=TLPQMjIwMzIwMjE5pJbUlgVkTA&index=3.

[28] Mohammed Jamjoon. What will break the stalemate in Myanmar? — Inside Story. Online broadcast. Al Jazeera English, March 5, 2021. https://www.youtube.com/watch?v=KH2eV5Dogxo.

[29] Roberts 2018. Putin also directed the police to monitor social media, including TikTok, in the aftermath of the Navalny protest videos; see Sherwin, Emily. Russia: Kremlin targets TikTok over content critical of Putin. Online. *DW News*, March 26, 2021. https://www.youtube.com/watch?v=DtrfP38OGy0.

[30] Tucker et al. 2017.

[31] Mohammed Jamjoom. What will break the stalemate in Myanmar? — Inside Story. Online broadcast. *Al Jazeera English*, March 5, 2021. https://www.youtube.com/watch?v=KH2eV5Dogxo.

[32] Paul Mozur, Mike Isaac, David E. Sanger, and Richard C. Paddock. "Facebook Takes a Side, Barring Myanmar Military After Coup". *The New York Times* (March 3, 2021). https://www.nytimes.com/2021/02/24/technology/facebook-myanmar-ban.html.

Chinese Digital Authoritarianism and Its Global Impact

Xiao Qiang, University of California at Berkeley[1]

The Rise of Chinese Digital Authoritarianism

A 2019 report by the Brookings Institution defines digital authoritarianism as "the use of digital information technology by authoritarian regimes to surveil, repress, and manipulate domestic and foreign populations."[2] Under Xi Jinping's leadership, the Chinese Communist Party (CCP) has been building digital authoritarianism in China through censorship, propaganda and AI-driven population-wide surveillance. Under the CCP's massive propaganda apparatus, a wide array of other organizations - including internet service providers, data analytics companies, and social media websites - also contribute to internet censorship and digital control of citizens in China. This comprehensive, full spectrum nature of the Chinese digital surveillance state is in many ways a goal for autocratic Arab regimes, one which the wealthier and high-capacity states such as the UAE have been actively pursuing. Understanding the Chinese model for digital authoritarianism can therefore shed light on potential futures of MENA states.

Since the early 1990s, China's Ministry of Public Security (MPS) has carried out the National Public Security Work Informational Project, also called the "Golden Shield Project."[3] This project includes a security management information system, a criminal information system, and a national adult citizen database (including fingerprints) among others. MPS has also established the Public Information Network Security Supervision Bureau to monitor, intercept and censor the online activities of Chinese citizens, from Bluetooth transmission to wireless networks.

In 2010, the penetration rate of Internet users reached about one-third of the 1.4 billion Chinese population. By 2016, more than half of the Chinese population was online. With the continuous growth of the number of Internet users in China, the Chinese government began to build the "Great Firewall" (GFW) in 2001, a collection of software and hardware systems used to monitor and filter communications on national internet gateways. The GFW surveils, intercepts, and blocks internet transmissions according to the official requirements of the CCP. It also blocks foreign internet tools and mobile apps, and forces foreign companies to adapt to domestic regulations.[4]

After Xi Jinping took office as the top leader of the Chinese Communist Party in 2012, he vigorously concentrated power by purging political opponents, promoting the CCP's ideology and his own personality cult, and strengthening the Party's complete control over society.

The Cyberspace Administration of China (CAC) is the central internet regulator, censor, oversight, and control agency for the Chinese government. The CAC answers to the Central Cyberspace Affairs Commission, which is headed by Xi himself. CAC has established a series of branches, including the Internet Commentary Work Bureau, the Mobile Network Management Bureau, the Cyber Security Coordination Bureau, and Internet Public Opinion Center.[5] Starting from 2018, CAC directly manages the National Computer Network and Information Security Management Center (aka the GFW).[6] (Prior to this, GFW was managed by the Ministry of Industry and Information Technology).

Growing Surveillance State

Another component of digital authoritarianism under Xi Jinping is an intensification of the mass retrieval, collection, and processing of individual information through online activities ranging from social media behavior to purchasing habits. This, combined with a vast and rapidly expanding constellation of cameras equipped with facial recognition systems, and crowdsourced reporting regimes, have enabled an unprecedented granularity of surveillance that enables individual behavioral manipulation. In the hands

of CCP, the new wave of digital technology is becoming a powerful, oppressive tool for surveillance and control of society as a whole.

In 2017, the Chinese government outlined its roadmap to become a "major AI innovation center in the world" by 2030.[7] The government selected Baidu, Tencent, e-commerce giant Alibaba and speech recognition software company iFLYTEK as the national champions in the AI field. These powerful companies are increasingly shaping reality as people's lives are becoming ever more dependent on their technologies—from intelligent voice assistants to various sensors, which collect data about people's living conditions and then analyze these data to improve people's quality of life.

The Chinese state works with tech companies to strengthen the large-scale retrieval, collection and processing of personal information through online activities ranging from social media behaviors to buying habits,.[8] With no other choice, more than one billion Chinese use a handful of phone applications. Although these phone applications are extremely convenient, users' communications, transactions, and behavior are disclosed to large technology companies such as Ant Group and Tencent that are obliged to share this data with the Chinese government.[9]

Chinese people have become accustomed to having their personal hobbies, education and health, academic qualifications, economic status, eating and consumption habits, social interactions, and even reading hobbies all in the vision of the big tech-companies and the state. This, combined with a large and rapidly expanding camera group equipped with a facial recognition system and a crowdsourced reporting system, achieves unprecedented monitoring granularity and allows individual behavior manipulation. For example, in China's north-west region Xinjiang, apart from the ubiquitous cameras, most residents are required to download apps on their phones that allow the authorities to monitor what they look at and track their movements. In 2019, data leaks revealed that Chinese authorities were closely tracking the locations of almost 2.6m people in real time through a facial-recognition company and police contractor called SenseNets.[10]

Skynet and Sharp Eyes

"Skynet Project" is a video surveillance project invested and established by the Chinese government in 2003. The government installs video surveillance equipment in public gathering places such as traffic junctions and security checkpoints, and uses GIS maps, image collection, transmission and other technologies to monitor and record information in different areas in real time. The Skynet system connects the surveillance cameras of different places (such as railway stations, restaurants, shopping malls, theaters and other public places, buses, subways, taxis and other transportation tools), and can identify a large number of people in a very short time. Chinese companies Hikvision, SenseTime, Huawei, and ZTE have all participated in the construction of the Skynet project. As of 2019, the Skynet system has 200 million public surveillance probes throughout mainland China.[11]

In 2015, China's National Development and Reform Commission (NDRC), the Central Political and Legal Affairs Commission, MPS, and six other government agencies launched the Sharp Eyes Project. This project's main goal is to provide complete real-time rural surveillance coverage by building high-definition cameras at main road entrances and crowd gathering places in rural areas. Sharp Eyes also places surveillance capabilities in citizens' hands and encourages their direct participation. The project uses existing rural TV networks to connect public safety video surveillance information to digital TV terminals of rural households. It aims to achieve "full range coverage, full network sharing, available at all times and fully controllable" from the perspective of police.[12]

Chinese authorities are thus integrating old and state-of-the-art technologies (phone scanners, facial recognition cameras, face and fingerprint databases, and many other technologies) into a wide range of tools for authoritarian control. For example, the "Skynet Project" face recognition system includes a series of application systems such as

a face capture and comparison system, a face retrieval system, and a video post-retrieval system.[13] These systems mainly use face detection algorithms, face tracking algorithms, face quality scoring algorithms, and face recognition algorithms, as well as personnel monitoring and motion tracking. Chinese police can access and search the face capture database into the face recognition server, and the server performs modeling and analysis on the face image. Chinese police can also deploy and install cameras at key monitoring locations to capture the faces of people passing through the Skynet.

Censorship, Propaganda and Disinformation

The CCP has always tried to enhance the legitimacy of the regime by shaping public discourse, mobilizing its support base and suppressing any political and social protest movements. The party-state proactively subverts and co-opts social media for their own purposes. Now it also uses algorithms, automation, and human curation to purposefully distribute misleading information to further enhance the effectiveness of its propaganda machine.

There are many types of "sensitive information" on the Chinese Internet, including so-called "internal information" such as propaganda prohibitions and inside stories of high-level political struggles, as well as social topics such as corruption, housing prices, medical reform, wages, and environmental pollution. The most censored topics include the Hong Kong protests, the anniversaries of the Tiananmen Square incident, and the detention of Uyghurs and other ethnic minorities in Xinjiang.

Mainstream applications such as Sina Weibo, Toutiao, and Kuaishou employ as many as thousands of people engaged in manual censorship to remove "illegal" content.[14] Many companies have outsourced content removal work to "censorship workshops" - a company called Beyondsoft has employed more than 8,000 workers. The Citizen Lab of the University of Toronto in Canada disclosed in August 2019 that WeChat already has image filtering capabilities. If users try to avoid text censorship and post sensitive content in images, they will also be discovered.

The CCP also uses high-tech censorship systems and official media reports, as well as social media platforms such as Weibo, WeChat, and LeTV, to increase internal and external ideological propaganda, especially manipulation of nationalism. The methods used include censoring information, distorting facts, changing narratives, and deliberately guiding people to forget history. As in the Gulf cases documented in this collection by Marc Owen Jones, a large number of social media accounts are supported or directly set up by government departments, and widely used deceptive digital tools, such as bots, botnets, and trolls.

For example, according to a *New York Times* report,[15] about 4,600 Twitter accounts reposted posts from Chinese diplomatic envoys and official news organizations in the first week of June 2020. During the Anti-Extradition Law Amendment Bill Movement in Hong Kong in 2019, Twitter suspended more than 200,000 fake accounts controlled by the Chinese government. These users have widely disseminated and created disputes about Hong Kong protests and deliberately "smeared the actions of Hong Kong demonstrators."

Pandemic

For China's rulers, the advance of censorship and surveillance technology can solve two fundamental problems: greatly reduce the cost of social coercion and suppression, and target the smallest resistance with the exact amount of force needed. COVID-19 has been a great gift to the digital authoritarian state.

The Covid-19 pandemic originated in China. One of the reasons why this pandemic spread so widely is directly related to the Chinese Communist Party's internet control. Before his death, Dr. Li Wenliang, known as the "whistleblower" of the Wuhan epidemic, was admonished by authorities for "spreading rumors." The death of Li Wenliang triggered great grief and anger among the Chinese people, and calls for freedom of speech flared online for a period afterwards.[16]

But this has not diminished the actual success of China's tight surveillance strategy. The Chinese government has been using various surveillance technologies during the COVID-19 pandemic, including tracking applications, drone surveillance, and cameras. Indoor and outdoor, remote temperature scanning and upgraded facial recognition can identify people wearing masks. A mobile application called "Health Code" has brought good news to public health and threats to privacy at the same time.[17] As people scan and board a bus or enter a restaurant, they can be stopped if they have a poor rating or have trouble using their smartphones. The "health code" has rapidly spread throughout China. Chinese people are very willing to enter their information into the health code program every morning.[18] The epidemic has become a long-term pass for the CCP's digital authoritarianism.

Digital Silk Road

Beijing's experience in using digital tools for home inspection and surveillance has made it the preferred supplier of illiberal governments wishing to deploy their own surveillance system. In the 65 countries assessed by the "Internet Freedom Report" by Freedom House in 2020,[19] Chinese officials have organized training courses and seminars for representatives from 36 countries in terms of new media and information management. Chinese state-owned and private companies are developing telecommunications infrastructure in 38 countries, and surveillance companies such as Hikvision and CloudWalk are selling facial recognition technology, using artificial intelligence, to 18 countries including Egypt and Qatar.

The Chinese government has been aggressively promoting its "Digital Silk Road" which is the code name for fiber optic cables, mobile networks, satellite relay stations, data centers and smart cities built by global Chinese technology companies. This effort has accumulated more than $17 billion in loans and investments, including funding for global telecom networks, e-commerce, mobile payment systems, and big data projects. China has specifically courted North Africa and the Middle East as part of its technology push; it reportedly has signed Digital Silk Road memoranda of understanding with Egypt, Saudi Arabia, Turkey, and the United Arab Emirates.[20]

Here are some latest examples from the "White Paper" published by official China Academy of Information and Communications Technology (CAICT) in April 2021:[21]

Alibaba has expanded its expansion plans to Southeast Asia. It has acquired Pakistani e-commerce company Daraz and launched a digital free trade zone with the support of the Malaysian and Thai governments, which will simplify customs inspections, provide logistical support for companies and promote exports of small and medium-sized companies in Malaysia and Thailand to China.

ZTE is currently operating in more than 50 of 64 countries on the route of the "One Belt, One Road" initiative. In addition to laying fiber optic cables and establishing mobile networks, the company has been providing surveillance, mapping, cloud storage and data analysis services to cities in Ethiopia, Nigeria, Laos, Sri Lanka, Sudan and Turkey.

Guan'an Information cooperated with the United Nations Asia-Pacific Region Economic and Information Technology Talent Training Center to establish the first domestic training base to provide professional safety training for countries along the "Belt and Road," with more than 200 annual training participants.

The Center for Strategic and International Studies (CSIS), a US think tank, issued a report on May 17, 2021,[22] stating that between 2006 and April 2021, Huawei had concluded 70 cloud infrastructure and e-government transactions with 41 governments or their state-owned enterprises. Most of these countries are classified by Freedom House as "non-free (34%)" or "partially free" (43%), concentrated in sub-Saharan Africa (36%) or Asia (20%), mostly low- and middle-income countries. Compared with advanced economies, these developing countries have "strong demand, lower barriers to entry and fewer scrutiny."[23]

Huawei processes a large amount of sensitive data related to citizens' health, taxation and legal records in the

contracting countries. Huawei Cloud Services also operates important infrastructure, such as oil production and fuel distribution in Brazil, and power plant operations in Saudi Arabia. This enables Chinese companies to collect, control, and store data from other countries, and access the data as needed. China is also very interested in analyzing huge amounts of data, trying to use it for artificial intelligence to help them improve their calculation and control models.

The Chinese government hopes that these companies can exert political influence throughout the region. In the short term, the arrival of Chinese engineers, managers and diplomats will strengthen the tendency of developing countries, especially those with authoritarian governments, to embrace the concept of China's closed Internet.

"The Great Digital Contest"

As Laura Rosenberger wrote in Foreign Affairs in 2020: "Democratic countries view information as an empowering force in the hands of people: the free and open flow of ideas, news, and opinion fuels deliberative democracy. Authoritarian systems see this model as a threat, viewing information as a danger to their regimes and something the state must control and shape."[24] Now the world is entering the era of artificial intelligence. As a technology that currently relies on the centralization of massive data, AI tends to empower centralized autocratic government rather than decentralized democratic governance.

China is already the richest, most powerful, and most technologically advanced dictatorship.

Using surveillance, censorship, and the manipulation of information, the Chinese Communist Party shores up its power at home while weakening democratic competitors abroad. In a Pew report published in 2020, a western Internet pioneer, technology developer and manager predicted: "By 2030, artificial intelligence-based surveillance systems that China will develop and export to the world will enslave 75% of the world's population. These systems will be 7 days a week. Every citizen is monitored 24 hours a day to monitor their every action."[25]

The rise and global expansion of Chinese digital authoritarianism is reshaping the balance of power between democracies and authoritarian states in what I call "The Great Digital Contest." China has provided the world with a blueprint for the establishment of a digital totalitarian state. The Middle East, as this collection shows, is already a central battlefield in this global struggle. All democratic states and civil society actors must work in solidarity to counter the global expansion of Chinese digital authoritarianism to defend and preserve freedom and dignity in the 21st century.

Endnotes

1 Research Scientist, Director of Counter-Power Lab, School of Information, University of California at Berkeley; Founder and Editor-in-Chief, China Digital Times, (https//chinadigitaltimes.net)

2 Alina Polyakova and Chris Meserole, "Exporting digital authoritarianism: The Russian and Chinese models," *Brookings*, August 2019, https://www.brookings.edu/research/exporting-digital-authoritarianism/

3 Sonali Chandel, Zang Jingji, Yu Yunnan, Sun Jingyao, and Zhang Zhipeng, "The Golden Shield Project of China: A Decade Later An in-depth study of the Great Firewall," *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, October 2019, https://www.researchgate.net/profile/Sonali-Chandel/publication/338361425_The_Golden_Shield_Project_of_China_A_Decade_Later-An_in-Depth_Study_of_the_Great_Firewall/links/5e136bce299bf10bc392fc09/The-Golden-Shield-Project-of-China-A-Decade-Later-An-in-Depth-Study-of-the-Great-Firewall.pdf

4 Richard Clayton, Steven J. Murdoch, and Robert N. M. Watson, "Ignoring the Great Firewall of China," In: Danezis G., Golle P. (eds) *Privacy Enhancing Technologies*. PET 2006. Lecture Notes in Computer Science, vol 4258. Springer, Berlin, Heidelberg, 2006, https://doi.org/10.1007/11957454_2

5 *See* the Official Website of China's Cyberspace Administration: http://english.www.gov.cn/

6 Rogier Creemers, Paul Triolo, Samm Sacks, Xiaomeng Lu, and Graham Webster, "China's Cyberspace Authorities Set to Gain Clout in Reorganization," *New America*, March 26, 2018, https://www.newamerica.org/cybersecurity-initiative/digichina/blog/chinas-cyberspace-authorities-set-gain-clout-reorganization/

[7] "AI Policy in China," *Future of Life*, April 2020, https://futureoflife.org/ai-policy-china/#:~:text=In%20July%202017%2C%20The%20State,leading%20AI%20power%20by%202030.

[8] Jamie P. Horsley, "How will China's privacy law apply to the Chinese state?" *Brookings*, January 29, 2021, https://www.brookings.edu/articles/how-will-chinas-privacy-law-apply-to-the-chinese-state/

[9] Julie Zhu, "Exclusive: Chinese regulators to push tech giants to share consumer credit data - sources," *Reuters*, January 11, 2021, https://www.reuters.com/article/us-china-big-tech-data-exclusive/exclusive-chinese-regulators-to-push-tech-giants-to-share-consumer-credit-data-sources-idUSKBN29G0M4

[10] Yuan Yang and Madhumita Murgia, "Data leak reveals China is tracking almost 2.6m people in Xinjiang," *Financial Times*, February 16, 2019, https://www.ft.com/content/9ed9362e-31f7-11e9-bb0c-42459962a812

[11] Coco Feng, "China the most surveilled nation? The US has the largest number of CCTV cameras per capita," *South China Morning Post*, December 9, 2019, https://www.scmp.com/tech/gear/article/3040974/china-most-surveilled-nation-us-has-largest-number-cctv-cameras-capita

[12] Dave Gershgorn, "China's 'Sharp Eyes' Program Aims to Surveil 100% of Public Space," *One Zero*, March 2, 2021, https://cset.georgetown.edu/article/chinas-sharp-eyes-program-aims-to-surveil-100-of-public-space/

[13] Jingchen Nie, "'Sharp Eyes" has applied face recognition technology in 16 provinces and cities across the country to boost security," *Xinhua News*, March 23, 2018, http://www.xinhuanet.com/legal/2018-03/23/c_1122578758.htm

[14] Li Yuan, "Learning China's Forbidden History, So They Can Censor It," *The New York Times*, January 2, 2019, https://www.nytimes.com/2019/01/02/business/china-internet-censor.html?module=inline

[15] Raymond Zhong, Aaron Krolik, Paul Mozur, Ronen Bergman and Edward Wong, "Behind China's Twitter Campaign, a Murky Supporting Chorus," *The New York Times*, June 8, 2020, https://www.nytimes.com/2020/06/08/technology/china-twitter-disinformation.html

[16] Verna Yu, "'Hero who told the truth': Chinese rage over coronavirus death of whistleblower doctor," *The Guardian*, February 7, 2020, https://www.theguardian.com/global-development/2020/feb/07/coronavirus-chinese-rage-death-whistleblower-doctor-li-wenliang

[17] Paul Mozur, Raymond Zhong and Aaron Krolik, "In Coronavirus Fight, China Gives Citizens a Color Code, With Red Flags," *The New York Times*, March 1, 2020, https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html

[18] Masha Borak, "China wants to keep health codes after the pandemic but users aren't so sure," *South China Morning Post*, June 3, 2020, https://www.scmp.com/abacus/tech/article/3087437/china-wants-keep-health-codes-after-pandemic-users-arent-so-sure

[19] Adrian Shahbaz and Allie Funk, "The Pandemic's Digital Shadow," *Freedom House*, Freedom on the Net 2020, https://freedomhouse.org/report/freedom-net/2020/pandemics-digital-shadow

[20] Joshua Kurlantzick, "China's Digital Silk Road Initiative: A Boon for Developing Countries or a Danger to Freedom?" *The Diplomat*, December 17, 2020, https://thediplomat.com/2020/12/chinas-digital-silk-road-initiative-a-boon-for-developing-countries-or-a-danger-to-freedom/

[21] "Chinese tech giant's global cloud strategy may give Beijing 'coercive leverage,'" *ANI*, March 29, 2021, https://www.aninews.in/news/world/asia/chinese-tech-giants-global-cloud-strategy-may-give-beijing-coercive-leverage20210529222803/

[22] Jin Qi, "Research: Huawei still getting contracts from developing countries" *Financial Times*, May 17, 2021, http://www.ftchinese.com/story/001092509?full=y&archive

[23] Ji Xi, "Huawei switch to the cloud, China competing with US in Asia, Africa and Latin America," *Voice of America*, May 21, 2021, https://www.voachinese.com/a/Huawei-cloud-coercive-leverage-20210520/5899020.html

[24] Laura Rosenberger, "Making Cyberspace Safe for Democracy: The New Landscape of Information Competition," *Foreign Affairs*, May-June 2020, https://www.foreignaffairs.com/articles/china/2020-04-13/making-cyberspace-safe-democracy

[25] Janna Anderson and Lee Rainie, "Many Tech Experts Say Digital Disruption Will Hurt Democracy," *Pew Research Center*, February 21, 2020, https://www.pewresearch.org/internet/2020/02/21/many-tech-experts-say-digital-disruption-will-hurt-democracy/

Transnational Digital Repression in the MENA Region

Marwa Fatafta, Access Now

"I have left my home, my family, and my job, and I am raising my voice. To do otherwise would betray those who languish in prison. I can speak when so many cannot."

- Jamal Khashoggi

Introduction

Since the Arab Spring, the internet in the Middle East and North Africa (MENA) region has morphed into a heavily policed and repressed space.[1] Alerted by its instrumental role in political organizing and exercising fundamental rights during the protests of 2011, Arab governments have taken a heavy-handed approach to internet regulation and governance. As they attempt to exert their territorial sovereignty and domestic control over a global digital sphere, new technologies have enabled them to extend their repression beyond their national borders in more convenient and cheaper ways.

State crackdown on exiled and diaspora activists is generally defined as transnational repression. It is not a new phenomenon, but it has become more entrenched and widespread globally over the last decade.2 Authoritarian regimes utilize their embassies and consulates as "satellite stations"3 from which they can intimidate and attack exiled activists, conduct surveillance of diaspora communities, limit and control their mobility by withholding consular services including renewing or issuing passports and official documents. Other repressive tactics include assassinations, rendition, forced disappearance, unlawful deportation, as well as the harassment, harm, and detention of relatives back in home country as a proxy punishment of exiled activists.

In the digital age, transnational repression has grown in scale and form. New technologies, such as surveillance technology and spyware, have ushered in unprecedented capabilities for repressive regimes to deter dissent abroad in more convenient and insidious ways. Governments are no longer restricted by diplomatic relations, intelligence sharing agreements, and networks of regime loyalists and informants to spy on the private communications, activities, and movement of their exiled targets. Nor, as Ahmed Shaheed and Benjamin Greenacre show in this collection, are they constrained by any binding global norms or international law governing the use of digital technologies.

The rise of diaspora activism after the Arab uprisings led the transnational repression of Arab regimes, both online and offline, to become more egregious and violent. Most notably in the cases of Egypt, Syria, and Saudi Arabia, human rights activists and political dissidents have fled their home countries in fear of imminent or potential reprisal. Many seek to survive the annihilation of civil society,4 and to be able to politically organize in situations where political activity is strictly prohibited in the home country. In September 2020, for instance, exiled Saudi activists launched the country's first opposition party, the National Assembly Party (NAAS), to push for democratic change in the Kingdom and to fight against the regime's relentless violence and repression.5 According to the United Nations High Commissioner for Refugees, at least 815 Saudi nationals applied for asylum in 2017, compared with 195 in 2012.6 The number is expected to reach 50,000 by 2030, according to the Saudi government's own estimates.7

This paper examines the recent escalation in transnational digital repression in the MENA region, and outlines three common tactics to crack down on dissent abroad: 1) the use of ambiguous and over-broad legislation for cross-border censorship; 2) the use of digital surveillance tools and private cyber mercenaries; and 3) the weaponization of social media platforms to censor, delegitimize, and intimidate activists, journalists, and regime critics.

1. Cross-border censorship and prosecution

Jordan and the UAE

More than half of Arab countries have enacted repressive cybercrime laws that undermine freedom of expression and authorize mass surveillance of internet users, while the others rely on existing legislation including counterterrorism laws to combat cybercrime, safeguard national security, and preserve public order and societal values.[8] Under such elastic and ambiguous terms, internet users in the region are routinely arrested, prosecuted, and imprisoned for innocuous Facebook posts and tweets. In the context of transnational repression, some of these laws have been applied to quell criticisms of foreign countries and censor information that are at odds with geopolitical alliances and state narratives in the region.[9]

In Jordan, internet users can be prosecuted under the Anti-Terrorism Law No. 55 of 2006, amended in 2014, for "disturbing relations with a foreign state."[10] This provision has been used to penalize Jordanian citizens who are critical of Gulf states and their monarchies on social media. In 2015, a senior Muslim Brotherhood official, Zaki Bani Rsheid, was tried before the State Security Court over a Facebook post in which he criticized the UAE and accused it of sponsoring terrorism, and consequently was sentenced to 18 months in prison.[11] Similarly, freelance journalist Jamal Ayoub was imprisoned in 2015 for writing an article criticizing Saudi Arabia's military operation in Yemen.[12] Most recently, on August 26, 2020, the Jordanian authorities arrested a well-known Jordanian cartoonist, Emad Hajjaj, for publishing a satirical cartoon mocking the normalization agreement between Israel and the UAE.

As a result, Jordanian citizens are wary of airing their opinion about Gulf states. According to an anonymous editor at state-run media outlet, the government actively discourages negative reporting or criticism of Gulf rulers.[13] This demonstrates the ability of the Gulf monarchies to co-opt neighboring countries and shape their domestic online media spaces through financial assistance and humanitarian aid.[14] The UAE is one of Jordan's biggest financial supporters. Together with Saudi Arabia and Kuwait, the three countries pledged an aid package of $2.5 billion to Jordan with over $1 billion deposited directly in Jordan's central bank.[15]

By the same token, the UAE has leveraged its cybercrime law, Decree Law No. 5 of 2012, to prosecute and imprison Jordanian nationals for criticizing their home government. In October 2020, a Jordanian resident of the UAE, Ahmed Etoum, was sentenced to 10 years in prison for criticizing the Jordanian government and the royal family in a Facebook post. Etoum was convicted of "deliberately [carrying out] an act against a foreign country that could damage political relations between the UAE and Jordan, by publishing on Facebook news and information that contain insults and ridicule toward Jordan, its king, and its government."[16] According to Human Rights Watch, one of the pieces of evidence used to incriminate and convict Etoum by the State Security Circuit at the Abu Dhabi Court of Appeals was "joining Facebook groups consisting of opponents of the Jordanian government abroad and posting comments ridiculing certain government decisions, reposting on his page government-issued news alongside comments claiming government corruption, and re-sharing online appeals by Jordanian citizens requesting social aid from the government."[17]

Etoum is not the only Jordanian national who has been imprisoned in the UAE. In 2015, journalist Tayseer Najjar was arrested and held in secret detention for two years, and in 2017, sentenced to three years in prison and a fine of a fine of 500,000 Dirhams ($136,000 USD) under the cybercrime law for posting content critical of the UAE's regional policies before he moved to the country.[18] Two Jordanian brothers were also detained in 2015, severely tortured, and sentenced to 10 years in prison and a fine of one million Emirati Dirhams for charges related to terrorism.[19]

2. Surveillance tech and cyber mercenaries

Saudi Arabia and the UAE

The murder of Saudi journalist Jamal Khashoggi, and the surveillance operation conducted prior to it, demonstrates

the extent to which repressive regimes are willing to go to silence their dissidents abroad.[20] An investigation by Citizen Lab at the University of Toronto revealed that Saudi Arabia had used Pegasus, a malicious mobile phone spyware produced by the Israeli company the NSO Group, to spy on Khashoggi's colleagues and associates including the phone of Saudi activist Omar Abdulaziz.[21] Abdulaziz, an ardent critic of the Saudi regime who is based in Montreal, was planning with Khashoggi a social media project called the 'Bee Army' to help combat pro-regime troll armies on social media. Other surveillance targets included a staffer at Amnesty International,[22] Saudi political satirist Ghanem Almasarir,[23] and a New York Times journalist, Ben Hubbard, who is known for his reporting on Saudi Arabia and the Crown Prince Mohammed Bin Salman (MBS).[24] Abdulaziz remains under threat. The Saudi authorities arrested two of his brothers and a number of his friends to pressure him into silence.[25] In June 2020, the Canadian police warned him of being a "potential target" of the Saudi regime with "credible information about a possible plan to harm him."[26]

The Saudi surveillance operation also included the recruitment of two Twitter employees in the company's headquarters in order to access private information of Saudi dissidents including their email addresses, phone numbers, and IP addresses. This operation led to the arrest of Abdulrahman al-Sadhan, a 37 year old aid worker, who was forcefully disappeared in 2018. In 2021, Al-Sadhan appeared before the Specialized Criminal Court in Riyadh and was sentenced to twenty years in prison, followed by a twenty-year travel ban, for his anonymous criticism of the Saudi authorities on Twitter.[27]

Despite the international outcry following Khashoggi's murder, and the call for an immediate moratorium on the sales, transfer, and use of surveillance technologies by UN human rights experts,[28] none of the perpetrators were held to account,[29] and the global surveillance industry continues to flourish at an estimated value of $12 billion.[30] The Gulf monarchies also continue to expand their surveillance capabilities. In June 2021, Israeli newspaper Haaretz revealed that MBS had acquired in 2019 a new zero-click hacking spyware, from another secretive Israeli company called Quadream.[31] These types of spyware are extremely malicious as they can automatically infect a target's device without any interaction from the target, such as opening an email or clicking on a link.

What is notably alarming is the aspiration of both Saudi Arabia and the UAE to become powerful regional tech hubs, which would allow them to harness further influence and domination in the region. The UAE has branched into developing its own home-grown surveillance technology by recruiting private digital mercenaries. In 2017, the Emirati cyber company, DarkMatter, was reported to have offered lucrative contracts to Israeli ex-intelligence officers working for the NSO Group with annual salaries as high as $1 million.[32] Prior to that, DarkMatter hired over a dozen former ex-NSA hackers for its clandestine surveillance operation 'Project Raven' to spy on foreign governments, militant groups, and human rights activists critical of the monarchy.[33]

The new UAE-Israel normalization deal, signed on September 15, 2020, is expected to further advance the UAE's capabilities on cybersecurity and surveillance fronts, evident by the meeting of the countries' cyber security chiefs in Tel Aviv directly after signing the agreement.[34] In 2020, the UAE hosted the Israeli cyber conference, Cybertech Global, in Dubai. They also launched a new cyber initiative named the "UAE-IL tech zone," which aims at bridging "technological, entrepreneurial, business, venture capital, and government collaborations between the UAE and Israel," and hopes to foster "in-depth personal and professional relationships and to continue building a stronger region through tech."[35]

3. Weaponization of social media platforms

Israel and Palestine

One well-documented tactic of transnational digital repression is the use of state-sanctioned troll armies to manipulate and steer conversations, turning social media into a battleground of narratives. Whereas citizens

and activists are using social media to criticize their governments, disseminate information, and document human rights abuses, governments are weaponizing those spaces to legitimize their own policies and to intimidate activists into silence through smear campaigns and online harassment.

In some jurisdictions, governments have set up Internet Referral Units (IRUs) whose mission is to monitor and detect 'harmful' or illegal content on social media. One prominent example of such units is the Israeli Cyber Unit established in 2015. Housed within the General Attorney's office, the Cyber Unit works closely with the Israeli police, defense and security agencies, as well as the Prime Minister's National Cyber Bureau to coordinate and tackle "crime and terrorism in cyberspace."[36] To do so, the Cyber Unit has set up an "alternative enforcement" mechanism to submit requests to social media companies for removal of individual content that violates the platform's terms of services. The Cyber Unit's requests do not follow a legal due process to determine the illegality of such content and safeguard the users' right to freedom of expression, who are not even aware that the Cyber Unit is requesting to censor their content.[37]

As such, none of these 'voluntary' government requests are covered by tech companies' transparency reporting. However, according to figures reported by the Israeli government, 95 percent of the Cyber Unit's requests are related to national security. 87 percent of the requests were made to Facebook, and 90 percent of them were actioned.[38] During the first 10 days of May, amidst the rising violence in Israel and the Occupied Palestinian Territories, the Israeli government had asked social media companies to delete more than 1,010 pieces of content. More than half of the requests were made to Facebook, and according to the Israeli government, Facebook took down 48 percent of them.[39]

In addition to the Cyber Unit, the government of Israel has sponsored and promoted a number of 'grassroots' initiatives to mass report content on social media. The Prime Minister's office recruited university students to engage in Hasbara activity on social media in exchange for full scholarships and financial payments.[40] Former Israeli intelligence officers also developed an application called Act-IL to carry out and coordinate campaigns where volunteers are directed to mass report certain content or boost others by liking and sharing them.[41] According to internal Facebook leaks, Israel was the top country in the world to report content under the company's rules for terrorism, with nearly 155,000 complaints in the week proceeding Israel's bombardment of Gaza on 11 May. It also came third in flagging content under Facebook's policies for incitement to violence and hate speech, "outstripping more populous countries like the US, India, and Brazil, with about 550,000 total user reports in that same time period."[42]

The role of social media companies in these struggles is a major new aspect of digital politics, as Joshua Tucker argues in his essay for this collection. While online platforms may not censor online speech as much as authoritarian governments would like them to, systematic cases of over-moderation, arbitrary takedowns and discrimination raise questions over the platforms' content moderation policies and their algorithms: who designs them and how?

In May 2021, Palestinian activists took to social media to protest against the forced eviction of families in the neighborhood of Sheikh Jarrah in East Jerusalem. Thousands of content were deleted on Instagram, Facebook, TikTok, and YouTube.[43] Facebook apologized for content takedowns citing a technical error. Nevertheless, further takedowns and restrictions ensued hindering users from uploading content, live streaming, sharing, liking, and commenting on posts. On Twitter, tens of accounts were also suspended, and many others restricted. Twitter's Legal, Policy and Trust & Safety Lead, Vijaya Gadde, explained during Access Now's human rights conference, RightsCon 2021, that these arbitrary suspensions were a result of their automated tools responsible for detecting spam. The algorithms are trained to detect behavior rather than content, so it falsely flagged and suspended abnormally active users in this period.

The egregious censorship of Palestinian content across different media platforms has heightened the need for transparency over how social media companies develop and implement their policies, as well as the lack of equality in enforcement of such policies. Facebook, in particular, has been over-moderating Palestinian content under a specialized set of platform policies since 2016, the year Israeli officials began mounting public and private pressure on social media companies to censor Palestinian content.[44]

One of Facebook's problematic policies is their policy on the term "Zionism," according to which it would remove attacks against Zionists when the term is used as a proxy for Jews or Israelis. The policy undermines freedom of expression in a number of ways. For one, it applies a narrow and singular worldview in which Jews and Israelis are made synonymous with Zionists, which would ultimately stifle legitimate political speech on Israel and Zionism. Secondly, it provides special protection to a political ideology, which Facebook typically does not classify as a protected group as compared to ethnic, religious, and gender groups. Thirdly, as there is no universal definition of hate speech, Facebook would need to provide considerable understanding of nuance and context to moderate this politically and historically complex word. But in order to moderate content at scale, Facebook entrusts its algorithms with this extremely sensitive task resulting in frequent and erroneous censorship.

A second problematic policy is how Facebook moderates the use of the Arabic word *shaheed* (martyr in English). Under Facebook's Dangerous Individuals and Organizations policy, which blacklists certain individuals and groups and actively removes content that supports or praises them, the use of the word *shaheed* can signal support, praise, and glorification of terrorism. The word *shaheed*, which comes from Islamic texts, is a widely-used expression among Arab and Muslim communities to describe individuals who were killed in conflicts (among other uses). So, how has Facebook arrived at this politicized interpretation?

In the Palestinian context, Facebook seems to have taken cues from the Israeli government despite its consistent denial of such influence. Israel considers Palestinian expressions such as '*Shaheed*' (martyr), '*Intifada*' (uprising), '*Sumud*' (stead-fastness) or '*Muqawamah*' (resistance) as terrorist or inciting terminologies. Take for instance the prosecution of the Palestinian poet Dareen Tatour. Tatour was arrested in October 2015 for publishing a poem on Facebook titled 'Resist, My People; Resist Them.' Her interrogation and prosecution were based on a Hebrew translation of the poem which referenced expressions that Tatour did not write. For instance, the word 'martyr' was translated to 'terrorist.' Two Palestinian children who were murdered by Israeli settlers and referenced in the poem, Ali Dawabsheh and Mohammed Abu Khdeir, were also described as 'terrorists.' As a result, Tatour was forcibly transferred to a settlement near Tel Aviv where she was placed under house arrest and banned from accessing the internet and receiving visitors. Eight months later, she was allowed to move back to her family but remained under house arrest. In July 2018, Tatour was sentenced to five months in prison, released two months later.

As evident by the Palestinian case, and similar cases in the MENA region, social media policies are often developed and shaped at the request or influence of governments, with the cooperation of social media companies, allowing them to tighten the noose around narratives of dissent and resistance online.

Conclusion

Since the Arab Spring, MENA governments have been adamant about closely monitoring and restricting what is said and shared online. And while activists have fled their countries to be able to speak and organize freely, authoritarian regimes have been able to extend their repression, aided by surveillance technologies and digital mercenaries, to crackdown on activists who are out of their physical reach. Social media platforms have also turned into 'war zones' in their own right, where governments actively try to censor online speech and intimidate activists through troll armies, internet referral units, and influence over platform's content moderation policies.

One important conclusion driven from the cases of transnational repression shared in this paper is that the encroaching digital authoritarianism in the MENA region should be studied and analyzed beyond the limitation of geographical borders and legal jurisdictions, especially in the context of counter-revolutions which have characterized the geopolitics of the region over the last decade. Transnational digital repression, together with the race to build and advance state cyber powers, serves a purpose beyond the immediate silencing of exiled activists. It aims to influence flows of data and information across the MENA region, shape, and control regional and global conversations, and ensure that any effort for democratization and regime change is actively thwarted and prematurely suppressed.

Endnotes

[1] Marwa Fatafta. 2021. "From Free Space To A Tool Of Oppression: What Happened To The Internet Since The Arab Spring?". Tahrir Institute For Middle East Policy. https://timep.org/commentary/analysis/from-free-space-to-a-tool-of-oppression-what-happened-to-the-internet-since-the-arab-spring/.

[2] Freedom House. 2021. "Out Of Sight, Not Out Of Reach: The Global Scale And Scope Of Transnational Repression". https://freedomhouse.org/report/transnational-repression.

[3] Dana M Moss. 2016. "Transnational Repression, Diaspora Mobilization, And The Case Of The Arab Spring". *Social Problems* 63 (4): 480-498. doi:10.1093/socpro/spw019.

[4] "Human Rights Council: Countries Should Take Bold Action". 2021. *Cairo Institute For Human Rights Studies*. https://cihrs.org/egypt-human-rights-council-countries-should-take-bold-action/?lang=en&ct=t(EMAIL_CAMPAIGN_COPY_02).

[5] The National Assembly Party NAAS. 2020. "Declaration Of The National Assembly Party (Saudi Arabia)". https://www.jadaliyya.com/Details/41767.

[6] Manal al-Sharif. 2019. "The Saudi Diaspora Of Dissidents In Exile Is Fighting Back". *The Washington Post*. https://www.washingtonpost.com/opinions/2019/09/30/saudi-diaspora-dissidents-exile-are-fighting-back/.

[7] "Saudi Arabia Still Trying To Coax Dissidents Home, Says Report". 2019. *Middle East Eye*. https://www.middleeasteye.net/news/saudi-arabia-still-trying-coax-dissidents-home-says-report.

[8] Yahya Shqair. 2019. "Cybercrime Laws In Arab Countries: Focus On Jordan, Egypt And The UAE". Arab Reporters for Investigative Journalism (ARIJ). https://en.arij.net/wp-content/uploads/sites/3/2019/12/Cyber-Crime-Laws-in-the-Arab-world-Policy-paper-by-ARIJ.pdf.

[9] Helmi Noman. 2019. "Internet Censorship And The Intraregional Geopolitical Conflicts In The Middle East And North Africa". *Berkman Klein Center Research Publication* 2019 (1). doi:10.2139/ssrn.3315708.

[10] Human Rights Watch. 2020. "Jordan: Release Prominent Cartoonist". https://www.hrw.org/news/2020/08/28/jordan-release-prominent-cartoonist.

[11] Human Rights Watch. 2015. "Jordan: 18 Months For Criticizing UAE". https://www.hrw.org/news/2015/02/19/jordan-18-months-criticizing-uae.

[12] International Press Institute (IPI). 2015. "Jordanian Journalist Detained Over Facebook Post". https://ifex.org/jordanian-journalist-detained-over-facebook-post/.

[13] Hincks, Joseph. 2020. "What The Arrest Of A Prominent Jordanian Cartoonist Says About The State Of Satire In The Arab World". The Time Magazine, , 2020. https://time.com/5887018/arrest-jordian-cartoonist-freedom-middle-east/.

[14] Deniz Gökalp. 2020. "The UAE'S Humanitarian Diplomacy: Claiming State Sovereignty, Regional Leverage And International Recognition". CMI Working Paper WP 2020:1. Chr. Michelsen Institute. https://www.cmi.no/publications/7169-the-uaes-humanitarian-diplomacy-claiming-state-sovereignty.

[15] Suleiman Al-Khalidi. 2018. "Gulf States Extend Over $1 Billion To Jordan As Part Of Aid Package". https://www.reuters.com/article/us-jordan-gulf-aid-idUSKCN1ME0WT.

[16] Human Rights Watch. 2020. "UAE: Jordanian Convicted For Criticizing Jordan On Facebook". https://www.hrw.org/news/2021/02/11/uae-jordanian-convicted-criticizing-jordan-facebook.

[17] Ibid.

[18] Human Rights Watch. 2018. "UAE: Release Imprisoned Jordanian Journalist". https://www.hrw.org/news/2018/12/20/uae-release-imprisoned-jordanian-journalist.

[19] "Jordanian Brothers Tortured And Sentenced To Ten Years In Prison In The UAE". 2019. *MENA Rights Group*. https://www.menarights.org/en/caseprofile/jordanian-brothers-tortured-and-sentenced-ten-years-prison-uae.

[20] Agnes Callamard. 2019. "Rep. Annex To The Report Of The Special Rapporteur On Extrajudicial, Summary Or Arbitrary Executions: Investigation Into The Unlawful Death Of Mr. Jamal Khashoggi". Human Rights Council.

[21] Bill Marczak, John Scott-Railton, Ron Deibert, Adam Senft, and Bahr Abdul Razzak. 2018. "The Kingdom Came To Canada: How Saudi-Linked Digital Espionage Reached Canadian Soil". The Citizen Lab. https://citizenlab.ca/2018/10/the-kingdom-came-to-canada-how-saudi-linked-digital-espionage-reached-canadian-soil/.

[22] "Amnesty International Among Targets Of NSO-Powered Campaign". 2018. https://www.amnesty.org/en/latest/research/2018/08/amnesty-international-among-targets-of-nso-powered-campaign/.

[23] Thomas Brewster. 2018. "Exclusive: Saudi Dissidents Hit With Stealth iPhone Spyware Before Khashoggi's Murder." https://www.forbes.com/sites/thomasbrewster/2018/11/21/exclusive-saudi-dissidents-hit-with-stealth-iphone-spyware-before-khashoggis-murder/#779bd2762e8b.

[24] Bill Marczak, Siena Anstis, Masashi Crete-Nishihata, John Scott-Railton, and Ron Deibert. 2020. "Stopping the Press: New York Times Journalist Targeted by Saudi-Linked Pegasus Spyware Operator". The Citizen Lab. https://citizenlab.ca/2020/01/stopping-the-press-new-york-times-journalist-targeted-by-saudi-linked-pegasus-spyware-operator/.

[25] Rosie Perper. 2018. "A Canadian Political Refugee Made Videos Criticizing Saudi Arabia — Now Saudi Authorities Have Arrested His Friends And Family". *Business Insider*. https://www.businessinsider.com/canadian-refugee-activist-family-arrested-saudi-arabia-feud-2018-8?r=US&IR=T.

[26] The Guardian. 2020. "Exclusive: Saudi Dissident Warned by Canadian Police He Is a Target". https://www.theguardian.com/world/2020/jun/21/exclusive-saudi-dissident-warned-by-canadian-police-he-is-a-target.

[27] Jacob Silverman. 2021. "Twitter Is Enabling Saudi Arabia's Brutal Crackdown on Dissent". https://newrepublic.com/article/161995/twitter-saudi-arabia-mbs-dissident.

[28] OHCHR. 2019. "UN Expert Calls For Immediate Moratorium On The Sale, Transfer And Use Of Surveillance Tools". https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=24736.

[29] "Two Years After Khashoggi'S Slaying, No Accountability For Spyware Firm Or Saudi Government". 2020. Blog. *Access Now*. https://www.accessnow.org/khashoggi-two-years-later/.

[30] Brad Smith. 2020. "A Moment Of Reckoning: The Need For A Strong And Global Cybersecurity Response". Microsoft On The Issues. https://blogs.microsoft.com/on-the-issues/2020/12/17/cyberattacks-cybersecurity-solarwinds-fireeye/.

[31] Gur Megiddo. 2021. "Secretive Israeli Cyber Firm Selling Spy-Tech To Saudi Arabia". *Haaretz.* https://www.haaretz.com/israel-news/tech-news/.premium.HIGHLIGHT-the-secret-israeli-cyber-firm-selling-spy-tech-to-saudia-arabia-1.9884403.

[32] Amitai Ziv. 2019. "Mysterious UAE Cyber Firm Luring Ex-Israeli Intel Officers With Astronomical Salaries". *Haaretz*. https://www.haaretz.com/israel-news/.premium-mysterious-uae-cyber-firm-luring-ex-israeli-intel-officers-with-astronomical-salaries-1.7991274.

[33] Christopher Bing, and Joel Schectman. 2019. "Exclusive: Ex-NSA Cyberspies Reveal How They Helped Hack Foes Of UAE". *Reuters*. https://www.reuters.com/investigates/special-report/usa-spying-raven/.

[34] Mohammed Soliman. 2021. "How Tech Is Cementing The UAE-Israel Alliance". *Middle East Institute*. https://www.mei.edu/publications/how-tech-cementing-uae-israel-alliance.

[35] 2021. *The UAE-IL Tech Zone*. https://www.uaeil.tech/about-us.

[36] "About The Cyber Unit". 2021. *The Office Of The State Attorney*. https://www.gov.il/en/departments/general/cyber-about#:~:text=The%20Cyber%20Unit%20engages%20in%20three%20major%20areas%3A&text=Conducting%20criminal%20files%20in%20the,at%20the%20Ministry%20of%20Justice.

[37] Adalah – The Legal Center for Arab Minority Rights in Israel. 2021. "Israeli Supreme Court Green Lights Israel's 'Cyber Unit' That Works With Social Media Giants To Censor User Content". https://www.adalah.org/en/content/view/10292.

[38] Netael Bandel. 2021. "Social Media Giants Deleted 159 Anti-Vaxxer Posts At Israeli Cyber Unit's Request". *Haaretz*, , 2021. https://www.haaretz.com/israel-news/.premium-social-media-giants-deleted-159-anti-vaxxer-posts-at-israeli-cyber-unit-s-request-1.9751943.

[39] "Hundreds Of Incitement Content Have Been Removed From Social Media". 2021. *Ynet*. https://www.ynet.co.il/digital/internet/article/SytSph7Fd.

[40] Barak Ravid. 2013. "Prime Minister's Office Recruiting Students To Wage Online Hasbara Battles". *Haaretz*. https://www.haaretz.com/.premium-social-media-hasbara-worth-millions-1.5320153.

[41] Ishmael N. Daro. 2018. "How An App Funded By Sheldon Adelson Is Covertly Influencing The Online Conversation About Israel". *Buzzfeed*. https://www.buzzfeednews.com/article/ishmaeldaro/act-il-social-media-astroturfing-israel-palestine.

[42] Ryan Mac. 2021. "Facebook Workers Accuse Company Of Bias Against Arabs And Muslims". *Buzzfeed*. https://www.buzzfeednews.com/article/ryanmac/facebook-employees-bias-arabs-muslims-palestine.

[43] Access Now. 2021. "Sheikh Jarrah: Facebook And Twitter Systematically Silencing Protests, Deleting Evidence". https://www.accessnow.org/sheikh-jarrah-facebook-and-twitter-systematically-silencing-protests-deleting-evidence/.

[44] The Guardian. 2016. "Facebook And Israel To Work To Monitor Posts That Incite Violence". https://www.theguardian.com/technology/2016/sep/12/facebook-israel-monitor-posts-incite-violence-social-media.

Social media manipulation in the MENA:
Inauthenticity, Inequality, and Insecurity

Andrew Leber, Harvard University[1] and Alexei Abrahams, University of Toronto[2]

Over the past decade across the Middle East, social media platforms have gone from being praised as 'liberation technologies' to being lambasted as tools of repression.3 Between 2009-2011, starting with Iran's 'Green Revolution' and continuing into the 'Twitter Revolutions' of the Arab Spring, social media platforms like Facebook and Twitter appeared to facilitate popular mobilization against authoritarians. Perhaps inevitably, however, such an incubator of unrest could not be left uncontested. Regimes that survived the Arab Spring, chief among them Saudi Arabia, subsequently invested substantial resources to manipulate social media discourse in their favor.4 Such top-down efforts moreover benefited from a growing climate of disillusionment over the Arab Spring, and a concomitant rise in counter-revolutionary mobilization.5 And while sometimes running afoul of platforms' terms of service,6 they have also drawn legitimacy in recent years from a growing extra-regional consensus over the prerogative of states vis a vis 'content moderation.'7

In this essay, we stress the ways in which centrally directed, technology-based platform manipulation by authoritarian regimes – i.e. bot armies controlled by security officers in Interior Ministries - are augmented or even outpaced by less centralized and more organic forms of manipulation. We classify existing literature on social media in the MENA within three broad trends of manipulation: 'inauthenticity', 'inequality', and 'insecurity' (summarized in Table 1).8 While the literature has touched on each of the nine 'buckets' in Table 1, we find it has dwelt predominantly on the first column of the table (centralized manipulation), and especially on the first row of that column (inauthentic activity). State officials certainly undertake all three forms of intervention, but so do pro-regime decentralized actors who, while products of a state-curated information environment, appear to operate somewhat independently from state command. Furthermore, social media's unequal and hierarchical nature - and the ability of regimes to target trendsetting "influencers" for co-optation or repression - has attracted less scholarly attention than the potential for 'bots' to simulate mass online behavior.

We conclude by encouraging a renewed research agenda that lends greater weight to these under-explored areas. This would entail deeper theorizing of the ends and means of manipulation, contextualization and comparison of pro-government campaigns, and the use of mixed-methods research designs that pair data analysis with ethnographic work and qualitative interviews that can explore the meanings of online activity as well as the multiple motivations for pro-government mobilization.

Inauthenticity

Studies of "digital authoritarianism" within the Arab Gulf monarchies have focused heavily on ways that inauthentic accounts (often state-backed) manipulate online narratives. In pioneering work, Marc Owen Jones has documented the role of bot armies in promoting sectarian rhetoric in Bahrain,9 simulating support for pro-Saudi comments by then-US President Trump,[10] and promoting anti-Qatar hashtags as "trending topics."[11] Other research has tied specific Saudi state actors to social media manipulation, especially during the Gulf Crisis - a multiyear standoff between Saudi Arabia, Bahrain, and the UAE, on the one hand, and Qatar on the other.[12] The ongoing civil war in Libya has also emerged as a major regional site of state-backed social media manipulation, involving bot networks affiliated with the UAE and Saudi Arabia as well as, to a lesser extent, Qatar and Turkey.[13]

Taken together, these studies challenge the idea that social media expressions represent "autonomous expressions of opinion by individuals,"[14] highlighting instead the role of centrally directed state messaging in shaping online

Table 1: The three "I's" of pro-government social media manipulation, with examples of centralized (state) activity, decentralized (non-state) activity, and activity that falls somewhere in between.

	Centralized ←	*Ambiguous*	→ **Decentralized**
Inauthenticity Use of digital automation or fabrication tools to give the impression of online popularity or authenticity.	"Bot" armies "Fake news" websites	Coordinated "support groups" of users State exploitation of `platform advertising features	Individual purchases of followers and "engagement"
Inequality Leveraging the outsize impact of "influencers" on social media discourse.	State officials building social media presence State pressures on platforms to censor content	State cooptation or coercion of pre-existing "influencers" Social media presence of state-regulated media	Independent users building social media presence through pro-regime rhetoric
Insecurity The use of online intimidation, physical-world repression and surveillance to curtail online activity - especially of "critical" influencers.	Direct repression Cybercrimes laws "Bot" mobs	Media & state officials encouraging social media mobs	Independent social media mobs "denouncing" or harassing other users

speech. Gulf governments' manipulation of online discussions has become a common trope in media coverage and analysis of regional social media, becoming almost the default explanation for pro-government narratives online.[15]

Still, not all unusual activity is inauthentic. Even gold-standard detection methods are prone to misidentify real users as fake,[16] and researchers may hold unrealistic assumptions about what authentic social media activity looks like.[17] Researchers should acknowledge that identifying bots thus remains an ongoing challenge, establish rather than assume the existence of bot networks, and reflect on how biases in bot identification methods inform interpretation of results.

Furthermore, the presence of bots does not necessarily imply state-backed operations. Private companies sell a wide range of automated engagement services, maintaining millions of fake accounts to that end – including in the Middle East and North Africa.[18] Even if states sometimes employ these firms,[19] the authors' own work has repeatedly noted incidents of isolated individuals with high levels of inauthentic support.[20] This often has more to do with celebrities or would-be influencers buying sizeable followings outright to boost perceptions of their popularity, though the same dynamics could apply in creating a perception of political support.[21]

To be sure, the absence of bot activity also does not rule out central coordination. Under Twitter's policies, for example, governments can purchase advertising tools that encourage "cultural customs and local protocols to show allegiance" – in other words, pro-government rhetoric.[22] Governments (as well as private firms they employ) might likewise pursue a hybrid approach incorporating both bots

and real users, utilizing "support groups" that promise real users more followers in exchange for promoting preferred messaging.[23] Still, we argue that researchers should not begin with the presumption of government-directed inauthentic activity.

Inequality

Approaches that account for the hierarchy of online discourse can highlight the extent to which state-backed social media manipulation incorporates real users, not only to evade platforms' anti-manipulation algorithms scanning for manipulation, but also to ensure that citizens engage with state narratives. Bots may flood social media sites to disrupt conversations or harass individual users.[24] Yet Saudi bot networks generate limited online engagement even compared with authentic Saudi Twitter users. Even prominent pro-government activity can exhibit quite limited bot activity.[25]

The clear leader/follower dynamics of social media platforms instead suggests we focus on accounts that have established themselves as "influencers," generating an outsize impact on online discussions. Playing to the cut and thrust of social media culture affords particular individuals (or the accounts they control) the symbolic capital necessary to mobilize displays of support online from loyal followers.[26] In analyzing #jamal_khashoggi (Arabic) over October-November 2018 on Twitter , for example, we found that just 50 such users - 0.07% of 69,595 accounts in our sample - garnered over *half* of all retweets[27] Control of these commanding heights of online rhetoric (directly or indirectly) would allow MENA regimes to send clear cues regarding the tone of permissible or desirable online speech, whether state actors seek to instill genuine loyalty[28] or merely the appearance thereof.[29]

The most influential accounts on social media often belong to people or organizations with substantial offline social capital -- movie or soccer stars, prominent religious figures, regional news agencies, wealthy elites, ministers of state, and so on.[30] In this regard, the 'offline' power advantages of the state carry over to the online space.

Beyond his control of various bot armies and "support groups," for example, Saud al-Qahtani (a Saudi royal court official who effectively served as the Kingdom's "media czar") maintained an active and open Twitter presence, earning glowing praise from state-regulated Saudi media.[31] While Qahtani's account was subsequently suspended from Twitter for manipulation practices, fellow royal-court advisor Turki Al al-Shaikh continues to maintain an equally expansive online presence.[32]

State officials enjoy an additional advantage over "ordinary" influencers in their ability to lobby platforms regarding content moderation decisions - hindering the ability of individuals to build followings by discussing "undesirable" topics. While both states and ordinary users can, in theory, request content takedowns, state officials have greater leverage in threatening not just a citizen-led boycott of a platform but a country-wide ban of the site (potentially denying platforms substantial market share).[33] Israel's 'cyber referral unit', for example, is a government agency dedicated to flagging social media content deemed problematic and prompting the relevant platform (Facebook, Twitter, YouTube, etc) to take action.[34] In May of this year, as mass protests erupted across historic Palestine, organizations like 7amleh and Access Now documented numerous incidents where pro-Palestinian content was censored for dubious reasons.[35]

Even if state actors enjoy considerable advantages in setting a pro-state agenda online, individual influencers can build online followings without clear state ties - either building on offline fame or becoming "self-made" influencers through online activity alone. Officials across the GCC have found that such social media "stars" can help convey desired messages to their respective publics.[36] While these influencers might be intrinsically motivated to make pro-government statements, authoritarian regimes can also deploy state resources to incentivize cooperation through promise of reward or threat of repression. Saudi officials reportedly considered both with respect to journalist Jamal Khashoggi, in the hopes that he might be "a profound addition to the [Saudi] Twitter army and cadre of government mouthpieces."[37]

While analysis of online social mobilization from below has typically focused on government critics, pro-government social mobilization is quite common within GCC social media communities.[38] In Saudi Arabia, this has often taken the form of intense nationalism, with users competing in displays of loyalty and attacks on perceived enemies of the Kingdom.[39] Saudi businessman Monther al Mubarak, for example, garnered a wide following through relentless attacks on Qatar, Islamists, and other perceived enemies of Saudi Arabia with the advent of the Gulf Crisis. Other nationalist accounts, such as @KSA24, remain anonymous.[40]

While it is difficult to tell how independent of the regime these pro-government voices are, evidence suggests that at least *some* users are relatively independent. Rival camps of Emirati and Saudi influencers engaged in days of recriminations over their countries' divergent policies toward the conflict in Yemen, even as official figures from each country sought to downplay the rift.[41] Furthermore, inauthentic accounts tied to Saudi social media firm *Smaat* tried to grab the attention of influencers such as Monther Al Mubarak through Tweets - which we would be unlikely to observe if both were directly on government payrolls.[42]

Understanding the role played by influencers is important as it suggests that authoritarian regimes do not need an overwhelming online presence in order to dominate online discourse. Even if MENA states work to establish "avenues for actors… to express [preformed grievances] against targets selected by the state,"[43] decentralized mobilization by loyalist influencers is less costly for regimes to sustain and harder for platforms to curtail with purely technical fixes. It also suggests that these tactics may backfire if some users turn their online influence back on the state itself. By the summer of 2020, for example, Saudi state television felt the need to air a pointed news segment about the dangers posed by Saudi Twitter accounts attacking fellow citizens in displays of excess patriotism.[44]

Insecurity

The emergent inequality of online political discourse implies not only a narrow clique of pro-regime influencers, but also a rarefied vanguard of opposition influencers. In recent weeks, for example, the hashtag #savesheikhjarrah trended in parallel with popular mobilizations in Jerusalem and across historic Palestine. While the hashtag drew worldwide engagement from hundreds of thousands of social media users, Twitter data suggest a mere handful enjoyed outsized attention.[45] Yet the visibility of these influential activists can in turn put them on the radar of security services, with targeted repression in turn clearing the way for pro-regime rhetoric to dominate online spaces.

For a start, most prominent critics openly identify themselves and have suffered repressive action against themselves and their families. During the #savesheikhjarrah protests, for example, Mona and Muhammad El Kurd, both residents of Sheikh Jarrah with wide followings on social media, emerged as influential narrators, live-tweeting events from the ground and accepting interviews with international media. They were subsequently harassed and even temporarily detained by Israeli police. Loujain AlHathloul, the Saudi women's rights activist recently released from jail, was a prominent social media influencer prior to her arrest.[46] Saudi dissident Omar Abdulaziz achieved fame through his YouTube channel, where he criticized the Saudi regime on camera. He now lives in exile in Canada, where a Saudi squad sent to assassinate him was fortuitously turned away at the border.[47] In his absence, his brothers in Saudi Arabia have been incarcerated.[48] Iyad el-Baghdadi, another prominent Saudi critic, identifies himself via his Twitter account and 'Arab Tyrant Manual' podcast. He lives in exile in Norway, where state security recently intervened to save him from assassination.[49]

Even when dissidents seek to maintain anonymity or coordinate privately, regimes find ways to surveil them.[50] In perhaps the most shocking example of this, Saudi Arabia, frustrated by anonymous dissidents on Twitter, recruited two moles inside of Twitter's San Francisco headquarters to access their data and de-anonymize them.[51] Even when such server-side compromise fails, however, an activist can be de-anonymized and surveilled by 'phishing' attacks, in which the activist is lured into clicking on a

link or authorizing a download that ultimately leads to the implantation of 'spyware' on their mobile device or personal computer. Amnesty International, for example, has documented the use of Germany-based FinFisher's 'FinSpy' spyware, purchased by Egyptian intelligence services, to compromise the devices of Egyptian activists.[52] As early as 2016, Citizen Lab caught Emirati authorities attempting to implant spyware on the iPhone of Emirati human rights agitator, Ahmed Mansour.[53]

The spyware, known as 'Pegasus', turned out to be the product of NSO Group, an Israeli technology company with close ties to Israeli military intelligence. Pegasus was later found to have been successfully deployed against dissidents and journalists in numerous countries worldwide, from Morocco[54] to Mexico,[55] effectively transforming their mobile phones into 24/7 digital informants. The range and hazard of such technology was underscored perhaps most dramatically in the summer of 2018, when a Saudi operator successfully deployed Pegasus on the iPhone of Saudi dissident Omar Abdulaziz, then residing in exile in Canada.[56] At the time, Abdulaziz was regularly using his iPhone to speak with Jamal Khashoggi, with whom he hoped to coordinate anti-regime activity on Twitter.[57] A few months later, Khashoggi was assassinated by a Saudi hit squad in Istanbul.

Regimes have also utilized more low-tech methods of identifying and silencing critics. Several Saudi activists and those close to them blamed a Saudi spying operation at Twitter for the arrest of several individuals inside Saudi Arabia, some of whom operated anonymous accounts critical of the government.[58] More directly, governments across the GCC have simply arrested activists who operate openly within their home countries, either warning them to dial back criticism or subjecting them to years-long legal proceedings and prison sentences.[59]

Jennifer Pan and Alexandra Siegel find that arresting online opinion leaders does not necessarily deter other would-be critics within Saudi Arabia, at least in 2011 to 2016.[60] Yet particularly within Saudi Arabia, the UAE and Bahrain, the past few years have seen a shift towards the unrelenting repression of all forms of opposition, where even voicing sympathy for detainees forms grounds for arrest. A year prior to his murder, Jamal Khashoggi lamented that any space for "loyal opposition" within Saudi Arabia has all but vanished.[61] Nor is this pattern isolated to the GCC. In Egypt, where the Sisi regime has violently repressed the Muslim Brotherhood since the summer of 2013, even liberal critics of the government (more likely to receive support from Western governments) have been repeatedly targeted for arrest.

Many critics had built up considerable online followings in years past when there was a slightly wider latitude for online speech. Yet new "cybercrimes" laws as well as tighter enforcement of existing laws likely deter anybody from following in their footsteps. Between 2006 and 2015, all of the GCC monarchies passed some form of cybercrimes legislation, typically with vaguely worded clauses that render practically any online statement a potential criminal offense.[62] Similar laws have also been used in Iraq to harass and intimidate opposition activists over social media postings and have recently been enacted in Egypt and Jordan.[63]

Such repression is typically the preserve of the state, yet even here pro-government citizens might augment regimes' repressive reach without being part of a formal security apparatus. Saud al-Qahtani encouraged such collaboration at the outset of the Qatar crisis by calling on his followers to add names to a "Black List" of those sympathizing with Qatar or criticizing the Kingdom.[64] Some nationalists in Saudi Arabia even appear to call for the arrest of those deemed "traitors to the nation" over and above what the actual authorities are concerned about, in one case trying (successfully) to get a local influencer jailed for noting that a local bakery was out of bread.[65]

Conclusion

With the bulk of existing research on social media manipulation focusing mainly on state-directed, inauthentic activity, our main recommendation is that researchers consider how these tools of manipulation interact with

online hierarchies as well as more straightforward repression of opposing viewpoints. We also encourage researchers to examine the ways that state strategies implicate "ordinary" citizens as collaborators in efforts to promote pro-government discourses and stifle criticism.

Understanding the complex chain that leads from state actions, through the tangle of social media networks and onward to citizens' perceptions and political actions, will require a very different skill set from API data wrangling alone. Researchers may benefit from a political economy lens that explicitly theorizes the motivations and goals of different actors to better establish what they hope to explain. Making sense of influencers' motives, and their relationship with state authorities, may require researchers to move offline to conduct interviews, mirroring similar efforts to make sense of US internet "trolls" and other influential users.[66]

Important efforts in this area are already underway. Marc Owen Jones moves beyond top-down inauthentic coordination to explore the role of private firms as brokers of deceit and catalogue his own extraordinary cat-and-mouse investigations of fake journalists and self-appointed regional experts who gain the ear of respected media outlets.[67] Mona Elswah and Mahsa Alimardani document the role of religious authorities in spreading medical misinformation across the MENA region during the covid-19 pandemic. Jennifer Pan and Alexandra Siegel assess whether visible repression does or does not echo through wider social networks - while being careful to note that these linkages themselves take place within broader contexts of repression.[68] Future work in this vein holds out promise that researchers will continue to explain, anticipate, and ultimately challenge the myriad efforts of authoritarian regimes to warp online discussions to their benefit.

Endnotes

[1] PhD Candidate, Department of Government, Harvard University, andrewmleber@g.harvard.edu.
[2] Postdoctoral fellow, Technology & Social Change Project, Shorenstein Center, Harvard University, alexei_abrahams@hks.harvard.edu.
[3] Larry Diamond, and Marc F. Plattner, eds. *Liberation technology: Social media and the struggle for democracy*.(Baltimore: JHU Press, 2012). Tufekci, Zeynep, and Christopher Wilson. "Social media and the decision to participate in political protest: Observations from Tahrir Square." Journal of communication 62, no. 2 (2012): 363-379. Bradshaw, Samantha, and Philip N. Howard. *The global disinformation order: 2019 global inventory of organised social media manipulation*. Project on Computational Propaganda, 2019. https://comprop.oii.ox.ac.uk/wpcontent/uploads/sites/93/2019/09/CyberTroop-Report19.pdf. Bradshaw, Samantha, and Philip N. Howard. "The global organization of social media disinformation campaigns." *Journal of International Affairs* 71, no. 1.5 (2018): 23-32.
[4] Marc Owen Jones. *Digital Authoritarianism in the Middle East: Deception, Disinformation and Social Media*, forthcoming manuscript (London: Hurst, 2021).
[5] Dalia F Fahmy., and Daanish Faruqi. *Egypt and the contradictions of liberalism: Illiberal intelligentsia and the future of Egyptian democracy*. Simon and Schuster, 2017. Kirkpatrick, David D. Into the hands of the soldiers: Freedom and chaos in Egypt and the Middle East. Penguin Books, 2019.
[6] "New disclosures to our archive of state-backed information operations," Twitter (blog post), December 20, 2019, https://blog.twitter.com/en_us/topics/company/2019/new-disclosures-to-our-archive-of-state-backed-information-operations.html
[7] See Mona El Swah, and Mahsa Alimardani. "Digital Apartheid: #SaveSheikhJarrah and Arabic Content Moderation". POMEPS Studies, 2021. See also the discussion in "Israel's "Cyber Unit" and Extra-legal Content Take-downs," *Lawfare* (podcast interview), April 29, 2021.https://www.lawfareblog.com/lawfare-podcast-israels-cyber-unit-and-extra-legal-content-take-downs
[8] We intend these three "I's" to complement Margaret Roberts' conceptualization of the three "F's" of censorship - fear (deterring citizens from criticism), friction (making it harder to access information), and flooding (providing too *much* information). While there is certainly some overlap, we focus more on how regimes mobilize pro-government opinions to monopolize online discourse versus only examining how critical internet use is prevented. *Censored: Distraction and Diversion Inside China's Great Firewall* (Princeton: Princeton University Press, 2018).
[9] Marc Owen Jones. "Automated sectarianism and pro-Saudi propaganda on Twitter," Exposing the Invisible, January 18, 2017, https://exposingtheinvisible.org/en/articles/automated-sectarianism/.
[10] Marc Owen Jones. "In Graphs: How pro-Saudi Twitter Bots Boost Donald Trump's Ego (and his retweet count)," November 13, 2017, https://marcowenjones.wordpress.com/2017/11/13/in-graphs-how-pro-saudi-twitter-bots-boost-donald-trumps-ego/
[11] Marc Owen Jones. "The gulf information war| propaganda, fake news, and fake trends: The weaponization of twitter bots in the gulf crisis." *International journal of communication* 13 (2019): 1400 to 1403.
[12] Nathan Patin (b33lzebub). "Lord Of The Flies: An Open-Source Investigation Into Saud Al-Qahtani," Bellingcat, June 26, 2019, https://www.bellingcat.com/news/mena/2019/06/26/lord-of-the-flies-an-open-source-investigation-into-saud-al-qahtani/.
[13] Mohamed Kassab and Andy Carvin. " A Twitter Hashtag Campaign in Libya: How Jingoism Went Viral," Medium (blog), June 6, 2019, https://medium.com/dfrlab/a-twitter-hashtag-campaign-in-libya-part-1-how-jingoism-went-viral-43d3812e8d3f; Kassab, Mohamed and Andy Carvin. "Libyan Hashtag Campaign Has Broader Designs: Trolling Qatar," Medium (blog), July 31, 2019,

https://medium.com/dfrlab/libyan-hashtag-campaign-has-broader-designs-trolling-qatar-8b2ba69c7334 ; Abrahams, Alexei and Joey Shea. "Coordinated Behavior in Libya's Regional Disinformation Conflict," *Lawfare* (blog), February 5, 2021. https://www.lawfareblog.com/coordinated-behavior-libyas-regional-disinformation-conflict

[14] Amaney A. Jamal, Robert O. Keohane, David Romney, and Dustin Tingley. "Anti-Americanism and anti-interventionism in Arabic Twitter discourses." *Perspectives on Politics* 13, no. 1 (2015): 56.

[15] See, for example, accounts of "Saudi Twitter" in Hubbard, Ben. *MBS: The rise to power of Mohammed Bin Salman*. Tim Duggan Books, 2020: 137-146; Hope, Bradley and Justin Scheck. *Blood and Oil: Mohammed bin Salman's Ruthless Quest for Global Power*, Hachette Books, 2020: 198:212; Jones, Rory. "In Saudi Arabia, Twitter Has Become a Tool to Crack Down on Dissent," *The Wall Street Journal*, November 7, 2019. https://www.wsj.com/articles/in-saudi-arabia-twitter-has-become-a-tool-to-crack-down-on-dissent-11573126932; Timberg, Craig and Sarah Dadouch. "When U.S. blamed Saudi crown prince for role in Khashoggi killing, fake Twitter accounts went to war," *The Washington Post*, March 2, 2021, https://www.washingtonpost.com/technology/2021/03/02/saudi-khashoggi-twitter-mbs/.

[16] Adrian Rauchfleisch and Jonas Kaiser. "The False Positive Problem of Automatic Bot Detection in Social Science Research," Berkman Klein Center Research Publication No. 2020-3, 2020 (revised February 11, 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3565233.

[17] See Marc Owen Jones, "Tracking Adversaries: The Evolution of Manipulation Tactics on Gulf Twitter", POMEPS Studies, 2021. Kazemi, Darius. Twitter Post. (June 3, 2020, 12:36 PM EST). https://twitter.com/tinysubversions/status/1268222451252518912

[18] Nicholas Confessore, Gabriel J. X. Dance, Richard Harris, and Mark Hansen. "The Follower Factory," *The New York Times*, January 27, 2018, https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html; Lieber, Chavie. "The Dirty Business of Buying Instagram Followers," *Vox*, September 11, 2014, https://www.vox.com/2014/9/11/7577585/buy-instagram-followers-bloggers.

[19] Shelby Grossman and Khadeja Ramani. "Outsourcing Disinformation," *Lawfare* (blog), December 13, 2020. https://www.lawfareblog.com/outsourcing-disinformation

[20] Andrew Leber and Alexei Abrahams, "Storm of Tweets," 253-254; Abrahams, Alexei, and Andrew Leber. "Comparative Approaches to Mis/Disinformation| Electronic Armies or Cyber Knights? The Sources of Pro-Authoritarian Discourse on Middle East Twitter." *International Journal of Communication* 15 (2021): 1184.

[21] Asir Ahmed. "هل يشترى المشاهير «متابعين» لحساباتهم على مواقع التواصل الاجتماعى؟ [Do influencers buy followers for their social-media accounts?]," *Youm7*, May 18, 2017, https://www.youm7.com/story/2017/5/18/هل-يشترى-المشاهير-متابعين-لحساباتهم-على-مواقع-التواصل-الاجتماعى-بالصور/3241109

[22] Marc Owen Jones. "Profit for Propaganda: Twitter Still Complicit in Whitewashing the Murder of Jamal Khashoggi," DAWN (blog), March 8, 2021. https://dawnmena.org/profit-for-propaganda-twitter-still-complicit-in-whitewashing-the-murder-of-jamal-khashoggi/

[23] Renée DiResta, Shelby Grossman, K.H., and Carly Miller. "Analysis of Twitter Takedown of State-Backed Operation Attributed to Saudi Arabian Digital Marketing Firm Smaat," Stanford Internet Observatory, December 22, 2019, https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/20191223_smaat.pdf; Grossman, Shelby and Khadeja Ramali. "Outsourcing Disinformation," *Lawfare* (blog), December 13, 2020, https://www.lawfareblog.com/outsourcing-disinformation.

[24] Roberts, *Censored*.; Angwin, Julia. "Cheap Tricks: The Low Cost of Internet Harassment," ProPublica, November 9, 2017, https://www.propublica.org/article/cheap-tricks-the-low-cost-of-internet-harassment.

[25] Christopher Barrie, and Alexandra Siegel. "Kingdom of Trolls? Influence Operations in the Saudi Twittersphere." *Journal of Quantitative Description: Digital Media* 1 (2021); Abrahams, Alexei, and Andrew Leber. "Comparative Approaches to Mis/Disinformation| Electronic Armies or Cyber Knights? The Sources of Pro-Authoritarian Discourse on Middle East Twitter." *International Journal of Communication* 15 (2021): 1173-1199.

[26] Susie Khamis, Lawrence Ang, and Raymond Welling. "Self-branding,'micro-celebrity'and the rise of Social Media Influencers." *Celebrity studies* 8, no. 2 (2017): 191-208; Rao, Venkatesh. "The Internet of Beefs," *Ribbonfarm* (blog), January 16, 2020. https://www.ribbonfarm.com/2020/01/16/the-internet-of-beefs/

[27] Including original tweets and retweets by other users. Abrahams, Alexei, and Andrew Leber. "Framing a murder: Twitter influencers and the Jamal Khashoggi incident." *Mediterranean Politics* (2020): 3-5. On similar patterns in the United States, see Schradie, Jen. "The digital activism gap: How class and costs shape online collective action." Social Problems 65, no. 1 (2018): 51-74.

[28] Calvert W Jones. *Bedouins into bourgeois: Remaking citizens for globalization*. Cambridge University Press, 2017: 12-36.

[29] Lisa Wedeen. "Acting" as if": symbolic politics and social control in Syria." *Comparative Studies in Society and History* 40, no. 3 (1998): 503-523.

[30] See Alexandra Siegel, "Official Foreign Influence Operations: International Broadcasters in the Arab Online", POMEPS Studies, 2021.

[31] "لهذه الأسباب.. القحطاني يتحدى «قذافي الخليج» ونجله جوعان
[For these reasons… Al-Qahtani challenges 'Qaddafi of the Gulf'," Okaz, September 7, 2019, https://www.okaz.com.sa/qatar-province-/na/1570271.

[32] See, for example, this satirical video he posted to his account. Turki AlAlShikh. Twitter Post. April 13, 2021 (1:07 pm EST), https://twitter.com/Turki_alalshikh/status/1382017633676066821.

[33] See examples in Vietnam, "Viet Nam: Tech giants complicit in industrial-scale repression," Amnesty, December 1, 2020, https://www.amnesty.org/en/latest/news/2020/12/viet-nam-tech-giants-complicit/; and India "India Covid: Anger as Twitter ordered to remove critical virus posts," BBC, April 26, 2021, https://www.bbc.com/news/world-asia-56883483.

[34] See also the discussion in "Israel's 'Cyber Unit.'"

[35] Alison E.M., Rahaf Carmel, and Sarah Abu Alrob Salahat. "Hashtag Palestine 2020," 7amleh –The Arab Center for Social Media Advancement, May 2021, https://7amleh.org//storage/Research%20and%20Position%20Papers/Hashtag_Palestine_English_2020_9May%20(1).pdf; Mnejja, Kassem and Marwa Fatafta. "Sheikh Jarrah: Facebook and Twitter systematically silencing protests, deleting evidence," accessnow, May 7, 2021, https://www.accessnow.org/sheikh-jarrah-facebook-and-twitter-systematically-silencing-protests-deleting-evidence/. See also El Swah, Mona, and Mahsa Alimardani. "Digital Apartheid: #SaveSheikhJarrah and Arabic Content Moderation". POMEPS Studies, 2021.

[36] Alanoud Alsharekh. "Social media and the struggle for authority in the GCC." *The Canadian Journal for Middle East Studies* 1, no. 2 (2016): 8-33.

[37] Hope and Scheck. *Blood and Oil*, 198-212

[38] This mirrors larger trends in the social movements literature - Grzegorz Ekiert and Elizabeth Perry recently offer a reminder that "modern states themselves organize citizens to act collectively in order to promote specific state goals and interests." "State-Mobilized Movements: A Research Agenda," in *Ruling by Other Means: State-Mobilized Movements*, eds. Grzegorz Ekiert, Elizabeth Perry, and Yan Xiaojun (Cambridge: 2020): 19.

[39] Imam Alhussein. "Saudi First: How hyper-nationalism is transforming Saudi Arabia," *Policy Brief*, European Council on Foreign Relations, June 2019. https://ecfr.eu/wp-content/uploads/saudi_first_how_hyper_nationalism_is_transforming_saudi_arabia.pdf

[40] Twitter account at: https://twitter.com/KSA24.
[41] See summary of exchanges by Leber, Andrew. "KSA-UAE: YEMEN, HADI AND THE STC (8/23-25)," *The Bitter Lake* (blog post), August 24, 2019, https://thebitterlake.blog/2019/08/24/ksa-uae-yemen-hadi-and-the-stc-8-23-24/; Leber, Andrew. "THE UAE AND ITS LEADERSHIP ARE A RED LINE, YET…," *The Bitter Lake*, August 28, 2019, https://thebitterlake.blog/2019/08/28/the-uae-and-its-leadership-are-a-red-line-yet/.
[42] DeRista et al, "Smaat," pg. 19.
[43] Ekiert and Perry, "State-Mobilized Movements," pg. 19.
[44] Al Ekhbariya. Twitter post (@alekhbariyatv), June 14, 2020 (10:12 am EST), https://twitter.com/alekhbariyatv/status/1272170003786276865.
[45] According to Twitter data we collected via the REST API, over 400k users worldwide tweeted the hashtag in the second week of May 2021, but just 1,933 garnered 80% of retweets, of whom the top 211 garnered 50% of retweets.
[46] Jasmine Badger. "Saudi Women Right-to-Drive Activists Deploy Twitter, Face Terrorism Court," *TIME*, February 6, 2015, https://time.com/3697073/saudi-arabia-women-drive-twitter/. For more on the arrests of Saudi critics and the consequences for online dissent, see Pan, Jennifer, and Alexandra A. Siegel. "How Saudi crackdowns fail to silence online dissent." *American Political Science Review* 114, no. 1 (2020): 109-125.
[47] Douglas Quan. "In the crosshairs of a crown prince? Canadian hit-squad claim just latest allegation against controversial Saudi royal," *Toronto Star*, February 13, 20121, https://www.thestar.com/news/canada/2021/02/13/in-the-crosshairs-of-a-crown-prince-canadian-hit-squad-claim-just-latest-allegation-against-controversial-saudi-royal.html.
[48] Tim Adams. "Khashoggi confidant Omar Abdulaziz: 'I'm worried about the safety of the people of Saudi Arabia'," *The Guardian*, February 20, 2021, https://www.theguardian.com/film/2021/feb/20/me-jamal-khashoggi-mohammed-bin-salman-omar-abdulaziz-the-dissident-netflix.
[49] "Iyad el-Baghdadi, activist in Norway, 'warned by CIA of Saudi threat'," BBC, May 19, 2021, https://www.bbc.com/news/world-middle-east-48199885.
[50] For an application to China, see Xu, Xu. "To Repress or to Co-opt? Authoritarian Control in the Age of Digital Surveillance." *American Journal of Political Science* 65, no. 2 (2021): 309-325..
[51] Ben Hubbard. "Why Spy on Twitter? For Saudi Arabia, It's the Town Square," *The New York Times*, November 7, 2019, https://www.nytimes.com/2019/11/07/world/middleeast/saudi-arabia-twitter-arrests.html.
[52] Amnesty International, "German-made FinSpy spyware found in Egypt, and Mac and Linux versions revealed", (2020), available from: https://www.amnesty.org/en/latest/research/2020/09/german-made-finspy-spyware-found-in-egypt-and-mac-and-linux-versions-revealed. Scott-Railton, Bill Marczak, Ramy Raoff, and Etienne Maynier, "Nile Phish - Large-Scale Phishing Campaign Targeting Egyptian Civil Society" (2017), available from: https://citizenlab.ca/2017/02/nilephish-report.
[53] Bill Marczak and John Scott-Railton, "The million dollar dissident: NSO group's iPhone zero-days used against a UAE human rights defender", The Citizen Lab (2016).
[54] Amnesty International, "Morroccan Journalist Targeted with Network Injection Attacks using NSO Group's Tools" (2020a), available from: https://www.amnesty.org/en/latest/research/2020/06/moroccan-journalist-targeted-with-network-injection-attacks-using-nso-groups-tools.
[55] John Scott-Railton, Bill Marczak, Siena Anstis, Bahr AbduRazzak, Masashi Crete-Nishihata, and Ron Deibert, "Reckless VII: Wife of Journalist Slain in Cartel-Linked Killing Targeted with NSO Group's Spyware" (2019), available from: https://citizenlab.ca/2019/03/nso-spyware-slain-journalists-wife
[56] Marczak and Scott-Railton (2016). Bill Marczak, John Scott-Railton, Adam Senft, Bahr AbduRazzak, and Ron Deibert, "The Kingdom Came to Canada - How Saudi-Linked Digital Espionage Reached Canadian Soil" (2018), available from: https://citizenlab.ca/2018/10/the-kingdom-came-to-canada-how-saudi-linked-digital-espionage-reached-canadian-soil.
[57] https://www.washingtonpost.com/world/secret-recordings-give-insight-into-saudi-attempt-to-silence-critics/2018/10/17/fb333378-ce49-11e8-ad0a-0e01efba3cc1_story.html
[58] Ryan Gallagher. "Spies in Silicon Valley: Twitter Breach Tied to Saudi Dissident Arrests," *Bloomberg*, August 19, 2020, https://www.bloomberg.com/news/articles/2020-08-19/twitter-security-breach-blamed-for-saudi-dissident-arrests.
[59] "Saudi Arabia: No Country for Bold Women," POMED, October 16, 2018, https://pomed.org/saudi-arabia-no-country-for-bold-women/; See database at "140 Characters: Online Activists Harassed and Jailed in Gulf Arab States," Human Rights Watch, https://features.hrw.org/features/HRW_2016_reports/140_Characters/index.html, accessed May 14, 2021.
[60] Pan and Siegel. " Saudi crackdowns."
[61] Jamal Khashoggi. "Saudi Arabia wasn't always this repressive. Now it's unbearable.," *The Washington Post*, September 18, 2017, https://www.washingtonpost.com/news/global-opinions/wp/2017/09/18/saudi-arabia-wasnt-always-this-repressive-now-its-unbearable/.
[62] Joyce Hakmeh. "Cybercrime Legislation in the GCC Countries," Research Paper, Chatham House, July 4, 2018, https://www.chathamhouse.org/2018/07/cybercrime-legislation-gcc-countries.
[63] "'We Might Call You in at Any Time' Free Speech Under Threat in Iraq," Human Rights Watch, June 15, 2020, https://www.hrw.org/report/2020/06/15/we-might-call-you-any-time/free-speech-under-threat-iraq; "How you will be affected by the new cybercrime law: A guide," *Mada Masr*, August 21, 2018, https://www.madamasr.com/en/2018/08/21/feature/politics/how-you-will-be-affected-by-the-new-cybercrime-law-a-guide/; Araz, Sevan. "Jordan adopts sweeping cybersecurity legislation," MEI, January 30, 2020, https://www.mei.edu/publications/jordan-adopts-sweeping-cybersecurity-legislation.
[64] Hubbard, *MBS*, 137-14
[65] "اعتقال نجم «سناب شات» أبو الفدا بسبب تصويره رف خبز فارغاً
[Arrest of SnapChat star Abu al-Fidda after he filmed an empty bread shelf," *al-Araby al-Jadeed*, June 3, 2020. https://www.alaraby.co.uk/%D8%A7%D8%B9%D8%AA%D9%82%D8%A7%D9%84-%D9%86%D8%AC%D9%85-%D8%B3%D8%B9%D9%88%D8%AF%D9%8A-%D8%A8%D8%B3%D8%A8%D8%A8-%D8%AA%D8%B5%D9%88%D9%8A%D8%B1%D9%87-%D8%B1%D9%81-%D8%AE%D8%A8%D8%B2-%D9%81%D8%A7%D8%B1%D8%BA%D8%A7%D9%8B
[66] Whitney Phillips. *This Is Why We Can't Have Nice Things: Mapping the Relationship between Online Trolling and Mainstream Culture*, MIT Press (2015).
[67] Marc Owen Jones. *Digital Authoritarianism in the Middle East.*
[68] Pan and Siegel. "Saudi crackdowns," 123.

Tracking Adversaries and First Responding to Disinfo Ops:
The Evolution of Deception and Manipulation Tactics on Gulf Twitter

Marc Owen Jones, Hamid Bin Khalifa University

Authoritarian regimes and other 'bad' actors in the Middle East are using social media for large scale deception operations. With little transparency from tech companies and poor regulation around disinformation, monitoring and tracking those operations falls uncomfortably upon journalists, activists and academics.[1] It is therefore necessary to share and discuss emerging techniques of identifying deception with academics across disciplines. It is also important to be transparent about detection methods in an environment where the terms 'bot' and 'troll' are frequently deployed against those who have opposing views. Being clear about methods of identifying deception can be instructive in a number of ways. Without identifying and acknowledging such deception, sociological studies of social media will inevitably be plagued with 'corrupted' data. Scholars using social media data must be adept at filtering out such deception.

Although studies of fake news and disinformation have usually focused on content, it is important to include the means of distribution (e.g. bots or sock puppets) of that content. For this reason, deception is the preferred term: 'Deception is the wilful manipulation of the information space through erroneous content and manipulated forms of distribution, with the intent to cause some form of harm through demonisation, adversarial content, omission, misdirection, whitewashing, or influencing information availability in the service of political power maintenance.'[2] Deception can often involve co-ordinated and inauthentic behaviour, which is why Facebook tends to label it CIB (Co-ordinated inauthentic behaviour).

Examining deception has other benefits. It can provide insights into social media governance and policy. Techniques of analysis can help us find fingerprints of certain deception actors, or in some cases, the entity behind such operations. With this in mind, this paper documents some of the evolving ways of identifying certain types of Twitter influence operations in order to highlight the scale, resilience and diversity of the phenomenon – particularly those involving sock puppets (a social media account purporting to be one person but operated by someone else) and bots (an automated account that posts content according to a computer script). In doing so, it also highlights a number of case studies that reflect the dominant tropes of MENA-focused influence operations.

Government, Big Tech and the Opaque Deception Assemblage

Governments around the globe, and particularly authoritarian regimes, have sought to use social media to manipulate domestic and international publics. As others in this volume have pointed out, the MENA region is rarely the object of focus when it comes to social media disinformation. Akin Unver notes, for example, an emphasis on USA-focused research.[3] In addition, Russia and China get a great deal of attention, partly reflecting the dominance of transatlantic security concerns in English-language scholarship. Despite this, countries in the MENA are some of the most active targets and perpetrators of influence operations and deception. This is especially true when it comes to the micro-blogging site Twitter.

Twitter has since 2018 publicized attempts to take down state-backed information operations (see DiResta, Goldstein and Grossman show in in this collection for discussion of Facebook takedowns).[4] Twitter takedowns are when Twitter removes accounts believed to be connected to state-backed influence operations. Since this time, of all the takedowns published, the bloc forming KSA, UAE and Egypt represent the second biggest actor in terms of number of accounts taken down. They are the most prolific global abusers of Twitter behind Russia

and China.[5] Iran also features in the top ten of Twitter information operation takedowns.

Technology companies such as Twitter should not be viewed as separate from the assemblage of forces that allow for the perpetuation and execution of influence operations. Social media companies have long been accused of neglecting marginalised communities in the developing world or siding with the authorities in human rights abusing states when it comes to content moderation and the censorship of resistance struggle. Ahmed Shaheed and Benjamin Greenacre discuss in further detail in this volume the accusation that Facebook have been accused of complying all too freely with Israeli requests to take down Palestinian content,[6] and more broadly of gravitating toward 'the richer, more powerful, and better-organized side.'[7] In the context of the Gulf, both Facebook and Twitter have been criticised for having their MENA operations stationed in the UAE. This has invited similar accusations of bias as well as concerns about human rights.[8] I made Twitter aware of regional manipulation as early as 2016,[9] when I provided a list of several thousand bot accounts promoting anti-Shia hate speech in Arabic. But the problem has persisted until at least 2020.[10]

Despite its move towards some transparency, Twitter's published data is a poor indicator of actual influence operations. Many potential state-backed information operations remain unregistered, so that the absence of data on Twitter's official archives does not mean other suspended accounts are not connected to a particular state actor. As a case in point, in 2020, Twitter released data for several accounts connected to a Saudi backed information operation against Qatar. However, the scale of the operation at the time was much bigger.[11] Thousands of suspicious accounts were detected by myself, and later suspended by Twitter, but were not released as part of the state-backed archives.[12] Similarly, while it is tempting to assume that political disinformation operations are primarily the domain of the state, multiple actors can be involved in this process, from individual 'hackers', to digital marketing firms. As is explored in this volume by Alexei Abrahams and Andrew Leber, influence operations can be conducted by multiple actors, from state to private individuals, and efforts need to be made to expand research 'beyond top-down efforts.'[13] While researchers of course hope to find so-called 'puppet-masters',[14] unambiguous attribution can be difficult, with limited opportunities for a smoking gun. Identifying deception is not the same as defining attribution. Crucially, though, deceptive influence operations and manipulated data can be identified without decisive attribution.

The first priority of disinformation is *detection*. Understanding the methods of manipulation are key to identifying influence operations. Without Twitter being fully transparent about what factors exactly lead them to determine what counts as state-backed information operations, academics, analysts and journalists have to be adept at reverse engineering or creating other means of determining suspicious activity in order to rapidly detect manipulation. What I call *first responding* to manipulation is imperative for achieving timely disruption of propaganda and disinformation. The longer such content lingers, the more potential it has for circulation and adversely impacting the information ecosystem.

The following questions are therefore key: What computational and non-computational techniques are being used to manipulate Twitter and promote authoritarian propaganda in the Persian Gulf? How is computational and-non computational manipulation evolving in the Persian Gulf and how can we detect it? How is it developing in terms of its ability to evade spam detection algorithms? What can such research tell us about Twitter governance in MENA? Is Twitter really doing enough to combat computational propaganda? What are some examples of trends manipulated using these methods? How can social media be made useful as a means of studying public opinion given all this manipulation?

Indicators of deception operations on Twitter

A common tactic of deception is the use of thousands of bots (automated Twitter accounts usually simulating real people) to promote specific propaganda or disinformation.

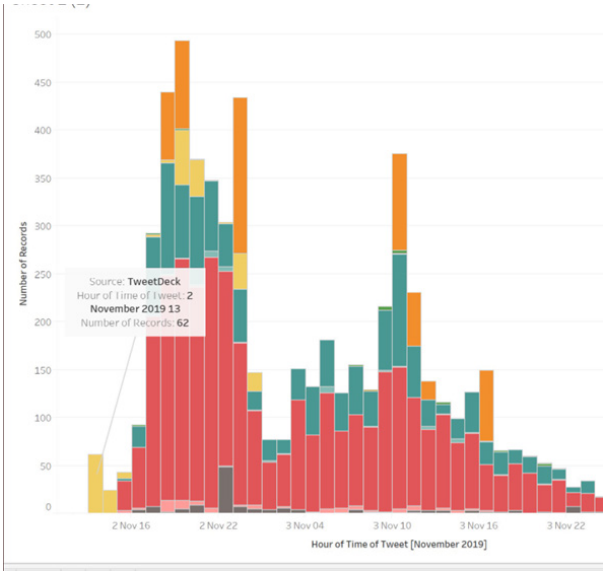
Deception here involves the content itself, but also creating the illusion of false consensus and imaginary publics (astroturfing). Twitter bots can be considered anomalies compared to organic Twitter activity. Since 2016, a useful method for anomaly detection has been looking at *unusual spikes in account creation*. Here, researchers can examine hashtags to see whether there are groups of accounts that are disproportionately created within a temporally limited time frame.[15] The logic here is that if hundreds or thousands of accounts are set up in a short time frame, and are also tweeting on the same topic, it can be said that there is a strong probability they were created solely for the purpose of platform manipulation. It used to be more common in the MENA region for bots to just copy and paste identical content across thousands of accounts. This seems less common now, but despite Twitter taking action to combat it,[16] this so-called 'copy-pasta' has not been wiped out completely.

On Gulf Twitter, at least, such indicators are perhaps less prevalent now than they were between 2016 and 2019. However, this does not mean inauthentic account creation is still not a relevant metric. As an example, in November 2019, thousands of accounts tweeting pro-Saudi propaganda around the Riyadh Agreement #اتفاق_الرياض were clearly created in a narrow time frame for the purpose of platform manipulation.[17] Many of the accounts engaged in this type of deception are connected to the Saudi-based news channel Saudi 24, which has been a prolific platform manipulator over the years.[18] This type of manipulation has been endemic. It has been used boost Donald Trump's anti-Iranian and pro-Saudi tweets, promote criticism of Qatar, spread anti-Shia hate speech, and promote praise of Mohammed bin Salman – to name but a few examples.[19]

Sometimes the *sequence of a trend and its initiation*, rather than account creation date, can be useful in detecting deception. As an example, in November 2019, the trend السكران_يدعو_السكارى_للاراده (The drunk calls the drunks to the square) began to trend in Kuwait. The 'drunk' is a reference to former Kuwaiti MP Saleh al Mulla and 'drunks' to those who support him. The trend appeared to be an attempt to deter people going to the square to protest. In this analysis, a number of other anomalies, including the application (e.g. was the Tweet sent from an iPhone or Android for example) used to send the Tweet, were arguably more useful than account creation in determining deceptive activity. Most strikingly, all of the accounts that seeded the hashtag used the same application – Tweetdeck; a highly unlikely organic scenario. This then prompted what seemed to be organic take up from real Kuwaiti Twitter users. A probable sock puppet farm then intervened; here accounts using the application Mobile Web (M2) app to exclusively retweeted a single account that was critical of Saleh al Mullasprung into action. Overall, low app diversity (a low number of unique applications) is an important signature, especially if those starting a hashtag appear to be using just one application (This may become less useful as a method of detection as Twitter has, over time, limited the existence of bespoke applications through stricter API access measures). It is also a good example of how a manufactured trend can then provoke genuine public discussion on a topic. This problematises the normative notion that the volume of bots is the issue. Indeed, if a small number of bots, trolls or influence 'operators' can successfully initiate and shape public discussion, then sheer volume of accounts is not always the most important thing in defining the narrative. This is highlighted well in Andrew Leber and Alexei Abraham's study of elite-driven narratives in the Gulf Crisis.[20]

Network analysis, a graph-based form of analysis showing the relationships between communities of accounts, can also be a useful means of detecting deception. In May 2020, Yemeni Nobel Prize winner Tawwokal Karman was appointed to Facebook's Oversight Council, a body created by Facebook to have the final say on key content moderation decisions.[21] Karman was perceived as a threat to some Gulf countries for her alleged support of the Muslim Brotherhood, and had been targeted with Karma spyware from the UAE.[22] Following the announcement from Facebook, online campaigns soon began to smear Karman, accusing her of being a terrorist, working for Turkey, or being an agent of the Muslim Brotherhood. English and Arabic hashtags trended, including "#no to Facebook Caliphate".

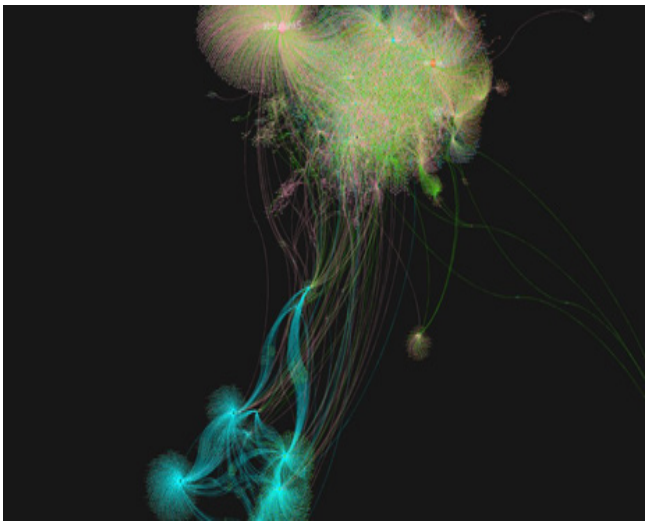
Figure 1 Tweets showing the hashtag 'The drunk calls the drunks to the square'. Yellow denotes tweets sent using Tweetdeck, Orange denotes M2



Network analysis of the hashtag 'no to Facebook Caliphate' demonstrated significant suspicious activity. The density of interactions, and the communities formed by interactions can indicate potential authentic versus inauthentic elements to those communities. In addition to creation date anomalies as mentioned before, the suspicious accounts formed distinct communities separate from denser clusters of communities. This relative isolation signified a lack of organic engagement with other users of the hashtag, which is itself unusual. The intra-community interaction between the distinct and separate community was high though, meaning these suspicious communities interacted a lot with one another, but not others - presumably to boost engagement and thus perceived popularity of the trend.

It was also clear that the already suspected community was sending tweets using the same Twitter application too. Many of the accounts were also created in a narrow time window. This series of compounding anomalies such as distinct modularity (community), low app diversity, and low creation date diversity indicate with even more certainty the existence of inauthentic and co-ordinated behaviour. The compounding anomalies also make it seem stranger that Twitter's algorithms did not flag the accounts

Figure 2 Network graph of no to Facebook Caliphate, showing suspicious community in bottom left quadrant



quickly. (In Figure 2, the blue accounts in the lower left quadrant indicate the separation of the suspicious community from the denser collection of accounts).

Chopped hashtags

An emerging trend in 2021 has been the use of '*chopped hashtags*'. Here, sockpuppet accounts dilute and pollute critical hashtags using abbreviated versions of the real hashtags.[23] So instead of, for example, #ماذا_استفدنا_من_الرؤية (what do we benefit from Saudi Vision 2030) sock puppets would deploy the hashtag #ماذا_ (what_). The impact seems to be that the chopped hashtags trend more readily than the ones they seek to replace. This tactic has been used multiple times to dilute trends critical of Mohammed bin Salman. This tactic became increasingly common and successful after the CIA released a report re-affirming the role of MBS in the murder of Washington Post journalist Jamal Khashoggi.[24] A similar tactic to this involves misspellings, such as khasxoggi instead of khashoggi.[25] It is thought this tactic exploits Twitter's trending algorithmic preference for novelty. (It is also worth noting activists may make use of misspellings in order to keep a topic trending.)

The dozens of instances of chopped hashtags since January 2021 generally share a common trait, one that likens them to some bot networks in some ways, but differentiates

them in others. Low centrality measures (how many times an account interacts with others or communicates) have become more useful than account creation date or intra-community interaction. That is to say, most of the accounts, when analysed, were not interacting with anyone but using the same suspicious hashtag seemingly on their own independent volition. The likelihood of multiple accounts independently deploying the same misspelled hashtag without some form of co-ordination is low. That many used the same application to tweet also created another compounding anomaly. Again, it is not clear who operates these accounts, but there is a wealth of accounts in Saudi that advertise paid trending services. Some even advertise services to get rid of 'unpatriotic trends', although they do not publicise their methods.[26] It has also been confirmed by a BBC investigation that unofficial Saudi-focused paid-for trending services work.[27]

Perpetrators: Not just state actors

Tracking manipulation, particularly when it involves anonymous accounts and bots, is also compounded by the difficulty in tracking down the perpetrators, who have different techniques of manipulation at their disposal. Where finding the perpetrators has been successful, it is clear that the number of actors involved highlights a diverse array of manipulators, each of whom could be selling their services or products to multiple clients - state or otherwise. In addition, it would make sense for state actors engaging in deception to potentially outsource projects in such a way as to remove obvious links to the state for the purposes of plausible deniability or operational security. In authoritarian regimes, the distance between state-linked accounts and private entities can be nebulous given the depth of autocratic control over multiple parts of the economy. For example, in Saudi Arabia, Twitter removed at least 88000 accounts connected to a digital marketing firm called SMAAT. One of their projects issued at 48 hours' notice was to cover the Riyadh Summit in 5 different languages.[28] Such a high-profile event is likely to involve some form of co-ordination with a government entity.

Where Twitter fails to offer clarity as to the provenance of certain manipulation operations, OSINT (Open Source Intelligence) and investigative work can yield results. In one instance, myself and Bill Marczak from Citizenlab managed to track down an Egyptian sock puppet operator through tracking metadata breadcrumbs. The manipulator eventually admitted to having created thousands of accounts that he managed through a software he created called Diavolo (Devil in Italian). They were mostly used to promote content for the channel Saudi 24 and its sister channels.[29] He later sold those accounts.[30] Similarly, by reverse searching a phone number included in an attempt to sell 5000 sock puppets to a Saudi-based individual via Twitter, I located a series of instructional videos for sockpuppet management software. The accounts that were run using this software would often include a tell-tale signature of a random string of alphanumeric characters, themselves designed (according to the creator) to fool Twitter's algorithm into finding the tweet unique and therefore not suspicious.[31] Accounts using this distinctive signature have been deployed on a number of networks, including one promoting ISIL propaganda to those tweeting on Saudi domestic politics.[32] Indeed, these random strings can also be useful indicators of sockpuppet activity. In both cases, there was no apparent limit to the number of accounts that could be operated by the software.

Beyond Bots

Bot has become a catchall term to describe a bad faith interlocutor, regardless of their authenticity. This is not necessarily a bad thing, as people tend to adopt such vocabularies as shorthand for general perceptions of manipulation. However, the ubiquity of the term should not conceal that there are many different things happening under the label of bot.[33] Deception operations go beyond sock puppets and bots. Co-ordinated tweeting (whether propaganda or disinformation) from a core network of influential accounts that then generates organic activity is also an increasingly common form of manipulating and controlling online discussions. A number of incidents in 2020, including viral rumours of a fake coup d'etat in

Qatar,[34] and the doxing of Al Jazeera journalist Ghada Oueiss with hacked personal photos, highlight how stories may be planted or orchestrated on social media, and then picked up by legacy media. Here influencers, sock puppets and bots generate buzz, with organic accounts and real media picking up the story. Because many of those involved are 'real' people, traditional markers of deception may not readily apply. However, such behaviour can still be considered co-ordinated manipulation, even if led by 'real' people with groups of sock-puppets and bots playing attendant and complimentary rolls. Tracking such campaigns requires significant contextual knowledge, ethnographic know how, and a broad array of digital tools.

Reading between the lines

What can the evolution of methods for detecting deception tell us about Twitter manipulation and indeed Twitter governance in MENA? There appears to be an evolution in the general techniques of computational and non-computational propaganda designed to evade detection by Twitter. A striking finding is the relative crudeness of such operations. It would be difficult to call many of the methods particularly sophisticated. The fact humans can identify such manipulation, call it out, only for it to continue, raises serious question about Twitter's will or competence in tackling it. That certain overt forms of crude manipulation can last for years is also a troubling reflection of Twitter's policy in the MENA region. The fact that Twitter CEO Jack Dorsey has met MBS twice, even after Twitter was compromised by spies acting on behalf of well-connected Saudi entities, has made observers suspicious about whether such tolerance is intentional. Similarly, while Twitter bans political advertising, it recently became apparent that advertising to express loyalty to political figures (including in authoritarian states) is permitted in Twitter's Terms and Conditions.[35] Without auditing and transparency, scholars should not place too much trust in social media companies' ability or intention to remain neutral arbiters of information wars. Certainly it would be remiss to rely on their data releases as a comprehensive source of manipulated content.

As social media becomes an important source and object of study, detecting social media deception is becoming an increasingly important skill for journalists, academics and analysts. Previous manipulations may have gone undetected if only one form of anomaly detection was applied. The potential lag in identifying deceptive content invites continued scrutiny and re-assessment of past scholarship. Many social media studies might be considered provisional, as the data used for such analysis might later turn out to be corrupted.

Although bots, and computational propaganda are important, it is important to see them as one component in a broader tapestry of deception. Trolls, bots, co-opted influences, all form part of a milieu that seeks to crowd oppositional voices out of the information space. Examining these holistically is perhaps more fruitful, even if much more challenging than anomaly detection. Certainly, the extent of deception is enough to warrant a critical re-evaluation of how we approach social media analysis. Indeed, we may be certain of what is false, but we cannot be certain of what is real.

Influence operations are increasingly more sophisticated, and locating them should not be reduced to counting the number of bots in a sample. We need more studies of deception (whether by bots or not) and influence operations to determine whether or not social media data can truly reflect authentic public opinion. The overarching ontology should be not to assume that social media is organic behaviour sullied by bots, but that no social media behaviour in the realm of MENA politics should be assumed to be organic.

Endnotes

[1] Claire Wardle. "10 questions to ask before covering misinformation". *First Draft*. 2017, https://firstdraftnews.org/articles/10-questions-newsrooms/
[2] Marc Owen Jones, Digital Authoritarianism, Deception, Disinformation and Social Media. (London: Hurst/OUP, 2021) (forthcoming)
[3] See Unver, A. Russian Digital Influence Operations in Turkey (this volume)
[4] Information Operations - Twitter Transparency Center. (2018). Twitter.com. https://transparency.twitter.com/en/reports/information-operations.html
[5] Marc Owen Jones. "Thought Russia was bad? Why Saudi Arabia is the world's most dangerous cyber bully". The New Arab. 2020, https://english.alaraby.co.uk/opinion/saudi-arabia-worlds-most-dangerous-cyber-bully
[6] Ahmed Shaheen and Benjamin Greenacre, Binary Threat: How Governments' Cyber Laws and Practice Undermine Human Rights in the MENA Region (this volume)
[7] Billy Perrigo. Inside Facebook's Meeting With Palestinian Officials Over Posts Inaccurately Flagged as Incitement to Violence. *Time. 21 May 2021.* https://time.com/6050350/palestinian-content-facebook/
[8] Marc Owen Jones "The Gulf Information War| Propaganda, Fake News, and Fake Trends: The Weaponization of Twitter Bots in the Gulf Crisis". International Journal Of Communication, 13, 27 (2019). Retrieved from https://ijoc.org/index.php/ijoc/article/view/8994/2604
[9] The author communicated with Twitter about several thousand Twitter accounts spreading sectarian hate speech. Twitter suspended around 1600 accounts, all of which showed the hall marks of temporally limited account creation
[10] Marc Owen Jones, "Automated sectarianism and pro-Saudi propaganda on Twitter". Exposingtheinvisible.org. Tactical Tech, (2016) https://exposingtheinvisible.org/en/articles/automated-sectarianism/
[11] Marc Owen Jones Anatomy of a disinformation campaign: The coup that never was. Aljazeera.com; Al Jazeera. (2020, May 19). https://www.aljazeera.com/features/2020/5/19/anatomy-of-a-disinformation-campaign-the-coup-that-never-was
[12] Marc Owen Jones (2020), May 16, Twitter Thread, https://twitter.com/marcowenjones/status/1261703612549521408
[13] Alexei Abrahams and Andrew Leber (2021) Social media manipulation in the MENA: Inauthenticity, Inequality, and Insecurity (in this volume)
[14] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, Alessandro Flammini Communications of the ACM, Vol. 59 No. 7 (July 2016,) Pages 96-104 10.1145/2818717
[15] See Jones "The Gulf Information War| Propaganda, Fake News, and Fake Trends:
[16] Twitter Comms, (2020), August, Status Update, https://twitter.com/TwitterComms/status/1298810494648688641
[17] https://twitter.com/marcowenjones/status/1193843347196133377
[18] Jones, "Automated sectarianism…"
[19] Jones, Digital Authoritarianism, 2021 (forthcoming)
[20] Andrew Leber and Alexei Abrahams "A Storm of Tweets: Social Media Manipulation During the Gulf Crisis". Review of Middle East Studies, 53(2), 241-258. doi:10.1017/rms.2019.45
[21] The New Arab. "Facebook appoints Yemeni Nobel laureate Tawakkol Karman as oversight board member". The New Arab. (2020). https://english.alaraby.co.uk/news/facebook-appoints-nobel-laureate-tawakkol-karman-oversight-board
[22] Joel Schectman and Christopher Bing Exclusive: UAE used cyber super-weapon to spy on iPhones of foes. U.S. (2019, January 30). https://www.reuters.com/article/uk-usa-spying-karma-exclusive-idUKKCN1PO19S
[23] Marc Owen Jones, "Profit for Propaganda: Twitter Still Complicit in Whitewashing the Murder of Jamal Khashoggi" – Democracy for the Arab World Now. (2021, March 8). DAWN. https://dawnmena.org/profit-for-propaganda-twitter-still-complicit-in-whitewashing-the-murder-of-jamal-khashoggi/
[24] See for example https://twitter.com/marcowenjones/status/1379806727667908608
[25] It is possible to that slightly misspelled hashtags are also used by activists to keep a topic trending. This is because's Twitter's algorithm rewards novelty, and hashtags will cease to trend as easily once they become commonplace.
[26] Marc Owen Jones (2021). March 23, Twitter Update, https://twitter.com/marcowenjones/status/1374302621234819076
[27] By BBC Trending. (2018, March 2). How much to fake a trend on Twitter? In one country, about £150. BBC News; BBC News. https://www.bbc.com/news/blogs-trending-43218939
[28] Smaat Co. LLC. (2018). Smaat.co. Website, https://smaat.co/en/projects/riyadh-summit
[29] Marc Owen Jones (2019). Oct 9 Twitter Thread, https://twitter.com/marcowenjones/status/1181972167963086849
[30] Yarno Ritzen. (2019, July 15). "How armies of fake accounts 'ruined' Twitter in the Middle East". Aljazeera.com; Al Jazeera. https://www.aljazeera.com/news/2019/7/15/how-armies-of-fake-accounts-ruined-twitter-in-the-middle-east
[31] Jones, Digital Authoritarianism….
[32] Marc Owen Jones Nov 2019, Twitter Status Update, https://twitter.com/marcowenjones/status/1200450510026919944
[33] Andrew Leber and Alexei Abrahams "Saudi Twitter blew up with support for the crown prince. How much of it is genuine?" Washington Post; The Washington Post. (2021, March 9). https://www.washingtonpost.com/politics/2021/03/09/saudi-twitter-blew-up-with-support-crown-prince-how-much-it-is-genuine/
[34] Jones, "Anatomy of a disinformation campaign…"
[35] Jones, "Profit for Propaganda…"

Follow the Money for Better Digital Rights in the Arab Region

Afef Abrougui, Independent Consultant and Researcher and Mohamad Najem, Executive Director, SMEX

Member countries of the Gulf Cooperation Council (GCC)—Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, and the United Arab Emirates (UAE)—are on a quest to untap the opportunities offered by technology to diversify their hydrocarbon-dependent economies. Sovereign wealth funds are investing in technology companies, tech giants and startups are incentivized to establish in the region, futuristic smart cities are under planning and governments are pouring money into improving and developing ICT infrastructure.

Such a business environment is conducive to both local and foreign tech startups and companies, and increasingly major players in the industry—including tech giants—are seeking to establish and operate in the region, despite potentially serious consequences for human rights. Digital authoritarianism is on the rise in the Gulf, as its autocratic monarchies deploy ever more sophisticated digital oppression tools and tactics to exert their power not only at home but also across the Arab region. Meanwhile international companies are turning a blind eye, favoring doing business in the region over human rights.

This paper is divided into three sections. In the first section, we explore what makes the GCC a conducive environment to the tech industry, and how the region's governments, particularly Saudi Arabia and the UAE, are exploiting technology to oppress populations and crack down on human rights and dissent. In the third and final section, we discuss the human rights implications of tech companies' operations in the region in the GCC's repressive regulatory environment.

Gulf investments in tech

The UAE has for years established itself as a tech hub in the GCC and the Middle East. Dubai is home to the MENA region offices of tech giants like Facebook, Google, and Twitter. Since its launch in 2007, the Telecommunications and Digital Government Regulatory Authority's ICT fund has been spearheading the development of the Information and Communication Technology (ICT) sector in the country by, for example, supporting and funding research and training in the field, in addition to incubator programs.[1] The current UAE cabinet includes a minister for advanced technology tasked with "enhancing the contributions of advanced sciences to the development of UAE and its economy," and a minister for artificial intelligence.[2]

As the push for economic diversification gears up in the region, other Gulf countries, and Saudi Arabia in particular, are trying to catch up.[3] With Crown Prince Mohamed Bin Salman's ascent to power, Saudi Arabia embarked on a series of reforms and programs to restructure and diversify its oil-dependent economy as part of its 2030 Vision.[4] For example, since the Vision's announcement in 2016, the kingdom increased its fiber optic capacity, introduced new technical programs on artificial intelligence and cybersecurity, launched a centre for the fourth industrial revolution with the World Economic Forum, and started a programme to transform Saudi oil giant Aramco into a leader in other sectors, including cloud services.[5]

A conducive environment for the tech industry

Multiple factors make the GCC a conducive environment to the tech industry. These include some of the highest internet penetration rates in the world, investments in infrastructure including fiber optic and 5G networks, and increased digitalisation.[6]

Investments from the region's sovereign wealth funds (SWFs), which has some of the largest funds in the world, are particularly attractive to technology companies and startups in the region and beyond.[7] For example, in 2020, both Saudi Arabia's Public Investment Fund (PIF) and Abu Dhabi's Mubadala Investment Company acquired

stakes in Indian technology company Jio Platforms, investing $1.5B and $1.2B, respectively.[8] The acquisitions are part of the funds' strategies to expand their ICT portfolios to contribute to economic diversification. PIF is also a shareholder in Uber after investing $3.5 billion in the American technology company that provides transportation and delivery services, including ride hailing and food delivery apps.[9] The investment earned PIF a seat in Uber's board of directors.

For its part, Mubadala has an ICT investment portfolio that includes social media apps and services, data centers, telecommunications and satellite operations.[10] Most recently, it announced a direct investment of $75m in encrypted messaging app Telegram, which is headquartered in Dubai. Abu Dhabi Catalyst Partners, which is jointly owned by Mubadala and the New York-based Falcon Edge Capital, invested an equal amount.[11]

The region is increasingly offering strong financial incentives such as tax breaks and low taxation rates for foreign and local businesses and investors. For example, as part of its 2040 Vision, Oman will exempt companies "in sectors aimed at economic diversification" from income taxes, if they start operating in the country between January 2021 and December 2022.[12] The region's multiple free economic zones offer similar incentives. Both Qatar and the UAE, have free trade zones exempting foreign companies from paying taxation and duties such as income and corporate taxes.[13] In these zones, foreign business owners and companies can also fully own their business and do not need a local partner. In some of these zones, tech startups are offered additional benefits. For example, innovative technology-driven startups in the Abu Dhabi Global Market (ADGM) zone can benefit from competitive license fees for five years and access to the accelerator programmes of HUB71, a global tech ecosystem in Abu Dhabi supporting startups to grow.[14] Under its Product Development Fund, Qatar Science and Technology Park (QTSP), a free zone hosting global tech companies, local startups and small to medium businesses offering "new high-tech products and services" can obtain grants that cover up to 50% of their total budgets.[15]

Tech as a lever of power

Beyond serving as an opportunity for economic growth and diversification, for GCC governments technology is used as a lever of power to control dissent and populations.

At home, control over the digital space is maintained through the use of spyware, internet filtering technologies, and trolls deployed to harass activists and dissidents and manipulate online discourses.[16] Technologies acquired from foreign companies are essential tools in the Gulf's digital oppression toolbox. For example, Saudi Arabia, the UAE, Qatar and Oman have all previously purchased surveillance systems from UK defence company BAE systems.[17] The purchases included Evident, an advanced tool developed by BAE's Danish subsidiary that enables governments to conduct mass surveillance of users' online activities, decrypt encrypted communications and determine the location of users based on data emitted by their mobile devices. In 2020, Israeli media reported that Bahrain, Oman, Saudi Arabia and the emirates of Abu Dhabi and Ras Al-Khaimah in the UAE signed contracts with the infamous Israeli NSO Group to acquire surveillance spyware.[18] The usage of NSO spyware has previously been documented in these four countries and Qatar, including for the purpose of spying on dissidents, journalists and human rights defenders.[19]

Recently, there have also been indications of an increased interest in tools and initiatives developed locally—with the support of foreign knowledge and skills in some cases.[20] Of GCC states, the UAE has emerged as a leader in leveraging its resources to deploy and promote local projects and technologies to help ensure its control over its citizens and residents online and offline. The most infamous of these initiatives is the privately-owned and UAE-headquartered spy firm DarkMatter, which bills itself as a cyber-security company.[21] DarkMatter is most notorious for its involvement in Project Raven, a spying campaign that targeted human rights defenders, critics of the Emirati government and other governments. [22] The company is also believed to be behind ToTok, a free messaging and video calling app released in 2019 and registered in the

Abu Dhabi Global Market economic free zone. The app quickly gained popularity in the Emirates, where the government has for years enforced a strict ban on most VoIP apps,[23] before security experts and technical analysis revealed it to be a spy tool[24] of the government capable of, among other things, tracking the conversations and images of its users.

Similar tactics are deployed in regional geopolitics. This particularly manifest during the GCC diplomatic crisis[25] of 2017, when Saudi Arabia, the UAE, Bahrain and Egypt cut diplomatic ties with Qatar and imposed an embargo on it for its support to Islamist groups, and specifically the Muslim Brotherhood, which is banned in Egypt, the UAE and Saudi Arabia, and its ties with Iran, a major regional rival for Saudi Arabia. Diplomatic ties with Qatar have since been restored, although experts warn that tensions remain.[26]

The feud played out online as the main players in the conflict, and the UAE and Saudi Arabia in particular, stepped up their usage of technology to target rivals and support their allies in the crisis through surveillance, cyber espionage[27] and online disinformation campaigns[28]. For instance, just two weeks before ties with Qatar were restored in January 2021, CitizenLab, an interdisciplinary lab at the University of Toronto that studies information controls, uncovered a hacking campaign targeting the iPhones of 36 journalists working at the influential Qatari-funded Aljazeera in July and August 2019.[29] The hackers, who the lab attributed to the UAE and Saudi Arabia, exploited a vulnerability in iMessage using NSO's Pegasus spyware.

Technology has also been aiding Gulf governments in their attempts to exert political influence in the wider Arab region and bolster autocratic regimes and rulers. A number of previously documented cases of social media disinformation campaigns in the region bear the fingerprints of the UAE and Saudi Arabia, and their allies. For example, in June 2019, days after the Sudanese military and the government-operated paramilitary group known as the Rapid Response Forces (RSF) cracked down on a pro-democracy sit-in demanding an end to military rule, massacring at least 100 people, a propaganda campaign praising the Sudanese military appeared on social media.[30] Experts believe the UAE and Egyptian governments were behind the campaign. Similar campaigns were waged in Libya, to support field marshal Khalifa Haftar, an ally of the Saudi-Emirati alliance and Egypt, in his attempt to overthrow a UN-recognized government.[31]

Human rights sidelined

Saudi-Emirati dominance over the GCC, poor human rights records, and investments in the tech industry, and specifically digital oppression tools, is bad news for human rights and democracy across the Arab region. This raises serious concerns about the exploitation of international platforms and services that receive investments from these two autocratic countries or establish themselves in this regulatory repressive environment where protections for freedom of expression and privacy are lacking.

In the aftermath of the 2011 Arab uprisings, as they became wary of mass protests sweeping through the region and toppling long serving regimes in Egypt, Libya, Tunisia and Yemen, GCC governments stepped up the legislative machinery to further tighten their control over the digital space. As Ahmed Shaheed and Benjamin Greenacre show in their contribution to this volume, the adoption of cybercrime laws that contain content-related offences criminalizing peaceful speech under vague terms and provisions proliferated.[32] For example, all GCC countries, with the exception of Bahrain, have provisions that criminalize and punish with imprisonment and fines prejudice to public order and morals in their cybercrime laws. In addition to fines, statements and calls to overthrow the regime or change the system are punished by up to 10 years in the Kuwaiti cybercrime law, 3 years in Qatar and an unspecified prison time in the UAE. Several other laws were also adopted. For example, in 2016, Bahrain enacted a law regulating newspapers in the digital space and requiring them to get permission from the authorities before disseminating news online.[33]

Citizens and residents are afforded little privacy protections due to lack of strong data protection laws and rampant government surveillance. Under Saudi Arabia's Cloud Computing Regulatory Framework, which was first introduced in 2018, Cloud Service Providers are required to "remove any Unlawful Content or Infringing Content from a Datacenter or other element of a Cloud System located in the Kingdom," and notify the authorities of any content "that may" violate the country's draconian cybercrime law.[34]

Business & Human Rights:

With the increased adoption of technology in the MENA region, especially the Gulf after the Arab Spring, international tech companies found the opportunity to expand and enter a new market. From a business perspective, this is a lucrative opportunity for these tech companies. For the monarchies in the Gulf, this is an opportunity to control the online space, and to build and improve their digital authoritarian empire. It was a win-win situation for companies and governments, but not for users in the Arab region, who are paying the price of international tech companies' profit-driven decisions to do business in the GCC. Many tech companies like Facebook and Twitter naturally gravitated towards the UAE, the most economically and technologically developed country in the Gulf, as well one of its most repressive.

With these new partnerships between the Gulf's authoritarian regimes and the tech companies came a price that normal citizens will pay. Tech companies like Facebook and Twitter claim that their platforms enable users to express themselves and exchange and access information. It wasn't until the end of 2015 that SMEX got its first contact with the MENA policy person in one of these companies. His work was mostly managing relationships with governments in the region. It was clear for our team that human rights are barely a second thought.

When it comes to freedom of expression, one of the interesting cases that was brought up to our attention in 2018 is related to Apple, when iTunes MENA refused to upload five songs of an underground Lebanese band named "Al Rahel al Kabir" had their album removed from iTunes MENA. Our team at SMEX did some investigation, and we discovered that there is a third-party company hired by Apple, called Qanawati, that took the decision not to upload the songs since they identified them as sensitive to our region.[35] The band was mocking ISIS leader, Baghdadi, and political oppression in the region. We did some campaigning and managed to get their music up through a Turkish third-party company, and the songs remained accessible on iTunes in the Gulf market.

In another example, Netflix censored an episode of comedian Hasan Minhaj's program *Patriot Act* because of a request from the Communications and Information Technology Commission (CITC) for breaching the cybercrime law in the Kingdom. The episode mocked Mohammad Bin Salman, the Kingdom's reaction to the disappearance of Jamal Khashoggi and the Saudi-led war in Yemen.[36]

Privacy and data protection are also under scrutiny in the companies' business operations in the region. Google announced a partnership with Aramco, a Saudi government-owned oil giant to start data centers inside Saudi Arabia which opens the door on collecting data from the whole region. Unfortunately, this is not the only project happening in the Gulf, with both Microsoft and Amazon on the same track.

Surveillance is a lucrative business opportunity for international tech spy firms and cyber security companies. For example, the UAE has been using Israeli NSO spyware to spy on its own citizens. In 2016, They tried to target Ahmed Mansoor, a prominent Emirati activist, with spyware exploiting an iOS vulnerability and capable of potentially hacking his phone.[37] The operations failed, and Apple released an update to close this gap. This scandal didn't stop UAE nor the NSO from doing more business in the region and target activists. It is believed that NSO and its spyware played a role in the killing of Jamal Khashoggi, the Washington Post columnist.38 Months prior to his assassination in October 2018, a successful surveillance

operation targeted Omar Abdel Aziz, another Saudi dissident living in Canada. The surveillance against Abdel Aziz exposed his WhatsApp conversations with Khashoggi about their potential plans for social media activism against the Kingdom. He believes the campaign played a key role in Khashoggi's killing inside the Saudi Consulate in Istanbul.

Tech companies are in bed with authoritarian regimes and dictatorships in the GCC, which represents a threat to human rights across the entire Arab region. As Western tech companies continue to expand their business operations in the Gulf, their tools and platforms are increasingly enabling these regimes to silence, surveil, torture and even kill their citizens. The implications for the digital space are far-reaching, particularly for the most vulnerable and at-risk communities including human rights defenders, journalists and dissidents who are at greater risks of surveillance, disinformation and harassment, and censorship and content takedowns.

Endnotes

1 'ICT Fund Introduction," The Telecommunications and Digital Government Regulatory Authority (TDRA) of the United Arab Emirates (UAE), Accessed July 8, 2021, https://www.tdra.gov.ae/ictfund/en/about-us/ict-fund-introduction.aspx.

2 ''Her Excellency Sarah Bint Youssef El Amiri," The Cabinet of the United Arab Emirates, Accessed July 8, 2021, https://uaecabinet.ae/en/details/cabinet-members/her-excellency-sarah-bint-yousif-al-amiri ; ''His Excellency Omar Bin Sultan Al Olama The Cabinet of the United Arab Emirates, The Cabinet of the United Arab Emirates, Acceded July 8, 2021, https://uaecabinet.ae/en/details/cabinet-members/his-excellency-omar-bin-sultan-al-olama.

3 Bahrain, Kuwait, Oman, Qatar, Saudi Arabia and the UAE have all launched long-term economic development plans emphasizing the role of technology, ICTs and the knowledge economy in diversifying their economies.

4 ''Vision 2030," Accessed July 20, 2021, https://www.vision2030.gov.sa/.

5 ''Vision 2030 Kingdom of Saudi Arabia," Accessed July 8, 2021, https://www.vision2030.gov.sa/ ; ''Progress and achievements," Vision 2030 Kingdom of Saudi Arabia, Accessed July 8, 2021, https://www.vision2030.gov.sa/v2030/achievements/ ; Amanda Russo, ''Building Emerging Technology Governance Key to Realizing Saudi Arabia's Vision 2030," World Economic Forum, November 6, 2019, https://www.weforum.org/press/2019/11/building-emerging-technology-governance-key-to-realizing-saudi-arabia-s-vision-2030/; Reuters Staff, ''Aramco to bring Google Cloud services to Saudi Arabia," Reuters, December 21, 2020, https://www.reuters.com/article/saudi-aramco-alphabet-idUSKBN28V1SB.

6 ICT", Arab Development Portal, November 2020, https://www.arabdevelopmentportal.com/indicator/ict-0 ; Rohma Sadaqat, ''Investments in ICT to drive digital economy," Zawya, September 1, 2020 https://www.zawya.com/mena/en/economy/story/Investments_in_ICT_to_drive_digital_economy-SNG_183509343/ ; Sophie Smith, ''Digital Transformation in the GCC," The Euro-Gulf Information Center, Accessed July 8, 2021, https://www.egic.info/digital-transformation-in-the-gcc.

7 ''Top 100 Largest Sovereign Wealth Fund Rankings by Total Assets," SWFI, Accessed July 8, 2021, https://www.swfinstitute.org/fund-rankings/sovereign-wealth-fund.

8 Mubadala Investment Company, Accessed July 8, 2021, https://www.mubadala.com/ ; ''Mubadala to Invest $1.2B USD in Jio Platforms," Mubadala Investment Company, June 4, 2020, https://www.mubadala.com/en/news/mubadala-invest-12b-usd-jio-platforms. ''Public Investment Funding Program 2021-2025," Public Investment Fund, Accessed July 8, 2021, https://www.pif.gov.sa/en/Pages/publicinvestmentfundprogram.aspx#6 ; Rishi Iyengar, 'Asia's richest man gets $1.5 billion from Saudi Arabia for his tech war chest," June, 19 2020, https://edition.cnn.com/2020/06/18/tech/mukesh-ambani-jio-saudi-investment-fund/index.htm.

9 ''Saudi Arabian Public Investment Fund Invests $3.5 Billion in Uber," Public Investment Fund, July 1, 2016, https://www.pif.gov.sa/en/MediaCenter/Pages/NewsDetails.aspx?NewsID=10.

10 ''Information & Communications Technology," Mubadala, Accessed July 8, 2021, https://www.mubadala.com/en/what-we-do/information-communications-technology.

11 ''Mubadala Investment Company, Mubadala and Abu Dhabi Catalyst Partners Invest $150 Million in Social Media Platform Telegram," Mubadala, March 23, 2021, https://www.mubadala.com/en/news/mubadala-and-abu-dhabi-catalyst-partners-invest-150-million-social-media-platform-telegram.

12 ''Oman: Incentives announced by the Government as part of Oman Vision 2040," PWC, March 14, 2021, https://www.pwc.com/m1/en/services/tax/me-tax-legal-news/2021/oman-incentives-announced-by-the-government-as-part-of-oman-vision-2040.html.

13 ''A complete guide to free zones in Dubai,' ; My Bayt, Accessed July 8, 2021, https://www.bayut.com/mybayut/guide-free-zones-dubai/ ; ''Frequently asked questions," Qatar Free Zone, Accessed July 8, 2021, https://qfz.gov.qa/resources/faq/.

14 'ADGM Tech Startup Commercial License," ADGM, Accessed July 8, 2021, https://www.adgm.com/faqs/adgm-tech-startup-commercial-licence#what-type-of-business-can-use-the-incentivised-tech-startup-licence ; ''Fostering Collaboration," HUB71, Accessed July 8, 2021, https://www.hub71.com/about-us/.

15 ''About Qatar Science & Technology Park," Qatar science & Technology Park, Accessed July 8, 2021, https://qstp.org.qa/about/ ; ''Funding SMEs and Startups that develop products and services relevant to the local market needs," Qatar science & Technology Park, Accessed July 8, 2021, https://qstp.org.qa/product-development-fund/.

[16] Carly Nyst and Nick Monaco, ''State Sponsored Trolling. How Governments Are Deploying Disinformation as Part of Broader Digital Harassment Campaigns," IFT, Accessed July 8, 2021, https://www.iftf.org/fileadmin/user_upload/images/DigIntel/IFTF_State_sponsored_trolling_report.pdf ; Jakub Dalek, Ron Deibert, Bill Marczak, Sarah McKune, Helmi Noman, Irene Poetranto, and Adam Senft, Tender Confirmed, Rights at Risk," Citizen Kab, September 21, 2016, https://citizenlab.ca/2016/09/tender-confirmed-rights-risk-verifying-netsweeper-bahrain/ ; ''The race to buy spyware in the Gulf," Fanack, January 26, 2019, https://fanack.com/role-of-the-gcc/spyware-gulf/.

[17] ''How BAE sold cyber-surveillance tools to Arab states," BBC, June 15, 2017, https://www.bbc.com/news/world-middle-east-40276568.

[18] TOI STAFF, ''Report: Israel pushed NSO spyware to Gulf states to help track dissidents," Times of Israel, August 23, 2020, https://www.timesofisrael.com/report-israel-pushed-nso-spyware-to-gulf-states-to-help-track-dissidents/.

[19] Bill Marczak, John Scott-Railton, Sarah McKune, Bahr Abdul Razzak, and Ron Deibert, ''Hide, and Seek. Tracking NSO Group's Pegasus Spyware to Operations in 45 Countries," Citizen Lab, September 18, 2018, https://citizenlab.ca/2018/09/hide-and-seek-tracking-nso-groups-pegasus-spyware-to-operations-in-45-countries/.

[20] For example, UAE spy firm DarkMatter, is known for hiring foreign expertise, including U.S. intelligence operatives.

[21] Darkmatter. Accessed July 8, 2021, https://www.darkmatter.ae.

[22] Christopher Bing, Joel Schectman, ''Special Report: Inside the UAE's secret hacking team of U.S. mercenaries," Jan 30, 2019, Reuters, https://www.reuters.com/article/us-usa-spying-raven-specialreport-idUSKCN1PO19O.

[23] Due to COVID-19, the UAE government started permitting a limited set of VoIP applications due to COVID-19, but access to other apps like Signal and Whatsapp is still restricted.

[24] Mark Mazzetti, Nicole Perlroth and Ronen Bergman''It Seemed Like a Popular Chat App. It's Secretly a Spy Tool," August 14, 2020, https://www.nytimes.com/2019/12/22/us/politics/totok-app-uae.html.

[25] ''Qatar crisis: Saudi Arabia and allies restore diplomatic ties with emirate," BBC, January 5, 2021, https://www.bbc.com/news/world-middle-east-55538792.

[26] Tarik M. Yousef, Omar H. Rahman, Noha Aboueldahab, Ranj Alaaldin, Adel Abdel Ghafar, Galip Dalay, Yasmina Abouzzohour, Robert P. Beschel Jr., and Nader Kabbani, ''What Brookings experts are saying about the breakthrough in the Gulf crisis," January 10, 2021, https://www.brookings.edu/opinions/what-brookings-experts-are-saying-about-the-breakthrough-in-the-gulf-crisis/.

[27] David D. Kirkpatrick and Azam Ahmed, ''Hacking a Prince, an Emir, and a Journalist to Impress a client," New York Times, August 31, 2018, https://www.nytimes.com/2018/08/31/world/middleeast/hacking-united-arab-emirates-nso-group.html.

[28] DFRLab, ''UAE Facebook pages targeted Qatar and Muslim Brotherhood," Medium, October 15, 2019, https://medium.com/dfrlab/uae-facebook-pages-targeted-qatar-and-muslim-brotherhood-8aec916fa1f7.

[29] [29] Bill Marczak, John Scott-Railton, Noura Al-Jizawi, Siena Anstis, and Ron Deibert, ''Journalists Hacked with Suspected NSO Group iMessage 'Zero-Click' Exploit," Citizen Lab, December 20, 2020, https://citizenlab.ca/2020/12/the-great-ipwn-journalists-hacked-with-suspected-nso-group-imessage-zero-click-exploit/.

[30] Declan Walsh and Nada Rashwan, ""'We're at War': A Covert Social Media Campaign Boosts Military Rulers," Citizen Lab, September 6, 2019, https://www.nytimes.com/2019/09/06/world/middleeast/sudan-social-media.html.

[31] ''In Libya, traditional and social media are used to fuel war," Arab Tyrant Manual, Accessed July 9, 2021, https://arabtyrantmanual.com/in-libya-traditional-and-social-media-are-used-to-fuel-war/.

[32] Joyce Hakmeh," Cybercrime Legislation in the GCC Countries,", Chatham House, July 4, 2018, https://www.chathamhouse.org/2018/07/cybercrime-legislation-gcc-countries/cybercrime-legislation-comparing-global-and-gcc.

[33] ''Decree 68 of 2016, regulating the use of newspapers for electronic media," Cyrilla, Accessed July 9, 2021, https://cyrilla.org/en/entity/szid4nc8pp29pweomkpj5rk9?searchTerm=bahrain.

[34] ''Cloud Computing Regulatory framework," Cyrilla, July 9, 2021, https://cyrilla.org/en/entity/9m28oc1kun5?searchTerm=Saudi%20Arabia&page=12.

[35] ''iTunes Yields To Petition From SMEX And Lebanese Band Al-Rahel Al-Kabir," SMEX, May 21, 2018, https://smex.org/itunes-yields-to-petition-from-smex-and-lebanese-band-al-rahel-al-kabir/.

[36] ''Netflix removes Hasan Minhaj comedy episode after Saudi demand," BBC, January 2, 2019, https://www.bbc.com/news/av/world-46740885.

[37] Bill Marczak and John Scott-Railton, ''The Million Dollar Dissident, NSO Group's iPhone Zero-Days used against a UAE Human Rights Defender," Citizen Lab, August 24, 2016, https://citizenlab.ca/2016/08/million-dollar-dissident-iphone-zero-day-nso-group-uae/.

[38] Access Now Team, ''Two years after Khashoggi's slaying, no accountability for spyware firm or Saudi government," Access Now, October 1, 2020, https://www.accessnow.org/khashoggi-two-years-later/.

Digital Orientalism:
#SaveSheikhJarrah and Arabic Content Moderation

Mahsa Alimardani and Mona Elswah, Oxford Internet Institute, University of Oxford

These voices will increase in number and volume. They cannot be ignored. (Mark Zuckerberg, 2012)[1]

Introduction

It is easy to forget in the current climate of scholarship centered on Silicon Valley disillusionment that social media companies were originally billed as conduits of revolution. The CEOs of these companies wanted the world to see them as tools that democratised information and access. The strategy of Facebook's Mark Zuckerberg during the company's stock market launch in 2012 was to take credit for pro-democracy movements like the 2011 Arab Spring.[2] Zuckerberg described the controlled media systems of countries living under censorship as *"the intermediaries controlled by a select few"* which would become liberated with the emancipatory features of his technology. A decade after the Arab Spring, the roles have been reversed. During the crisis surrounding Israeli settlers seizing homes in the Sheikh Jarrah neighborhood of Jerusalem, the Palestinian movement against what B'Tselem and Human Rights Watch have recently called Israeli apartheid found the policies developed by Facebook, and to a lesser extent other social media companies, a major obstacle to their mobilization.[3]

In this paper, we zero in on Arabic content moderation. We identify the systemic policies that are being administered by social media companies, whether designed within the technology or implemented through policies. While there have been issues with other platforms, we focus our analysis and argument in this paper on Facebook as the most egregious violator, with far reaching systemic problems and impact on Arab and pro-Palestine content. Second, we identify the various formats of digital repression of speech regarding the rights movement supporting Palestinians online. Third, we look at the countermeasures users have employed to overcome this repression as part of a greater movement for accountability from platforms in the Arab region. In this paper, we contend that failures and subjectivity of platform governance has given rise to what we call a new orientalism in the digital sphere, or digital orientalism. Orientalism is the stereotypical and discriminatory lens by which western nations view the Middle East and North African region. Western countries have used this lens to assert dominance and colonialism, either through war, media, governance and policies. We argue that this framework now defines the policies and actions Western social media companies use to disadvantage Internet users in this region.

A History of Facebook's Problems with Content Moderation

While the issues in the region have been playing out across many platforms, Facebook retains a particularly significant hold over online communications across the Arab world, with millions of users scattered across their three platforms: Facebook, Instagram and WhatsApp. Its influence on communications and media has come to the fore globally, especially since the 2016 United States Presidential elections[4] and the use of Facebook in scandals such as Cambridge Analytica. This increase in focus and concern has been exacerbated by studies that have outlined Facebook's outsized role in inciting a genocide by the Myanmar military against the Rohingya Muslims in 2018.[5] Due to flaws in the architectural design of Facebook, some scholars have questioned Facebook's legal responsibility and whether the company could be held accountable for its roles in these scandals.

Here we focus on deep rooted problems inherent specifically in Facebook's content moderation policies, which are ad hoc and inconsistent. In 2018, the *New York Times* published leaked content moderation guidelines

and practices through troves of PowerPoint presentations. They concluded that "the Facebook guidelines do not look like a handbook for regulating global politics. They consist of dozens of unorganized PowerPoint presentations and Excel spreadsheets with bureaucratic titles.[6] The investigation revealed that policies are designed for moderators to use Google Translate, as Facebook remains short on moderators who speak local languages. They rely on translations that often miss the nuances or facts of the context of speech at hand, a major problem in a region where Arabic is spoken with diversity in dialect and cultures. Although the company has taken some measures to prevent further scandals, Facebook's content moderation has continued to impose harm — especially in the Arab world.

The Hurdles of Arabic Content Moderation

Arab activists have been part of the broader digital rights movement calling out commercially-oriented social media platforms for their problematic positions. Since 2011, the policies and teams concerned with the Middle East and North Africa have developed in response to urgent pressure by users, governments, issues, and events. Prior to the Arab Spring, issues of content moderation were only resolved for elite and well-connected users. The most famous case is of Facebook's removal of the popular "We Are All Khaled Saeed" page[7] for going against its "real name policy"[8] prior to the start of the 25 January 2011 Egyptian Revolution. The page was restored only because of the connections of Wael Ghonim, one of the anonymous administrators of the page, who worked for Google and used his contacts to get in touch with Facebook's Chief Operating Officer Sheryl Sandberg to revive the page.[9]

The issues within the region have been persistent over the past ten years (see Table 1). We identify five forms of platform bias. First, the removal of pro-democracy Arabic content (e.g., posts, tweets, pages) has harmed many activists in the region. Second, Arab activists have repeatedly had their accounts restricted and deleted on the basis of violating the platform's community standards. In Arab countries, many pro-democracy pages, groups, accounts, and content have been taken down, with their accounts suspended or de-platformed for what companies would call "Terrorist and Violent Extremist Content" (TVEC), hate speech, organized hate, hateful conduct, and violent threats. Second, even unintentional removals through automated systems have far reaching consequences. For example, YouTube's community guidelines prevent publishing graphic and violent videos which have mistakenly led their algorithms to take down several videos from Syria that documented the war crimes of the regime of Bashar al-Assad.[10] From 2012 to 2019, YouTube erased about 206,077 videos related to Syria and removed several channels owned by activists and local news outlets.[11]

Third, the Global South, including the MENA region, faces double standards compared to the rest of the world. For example, activists and researchers have noted the limited access to social media data during elections and other political crises.[12] This was evident during Tunisia's 2019 elections when civil society members could not benefit from the archives of political ads in the Facebook Ad library.[13] While researchers and activists in many countries were able to monitor political ads and know which audiences are being targeted by politicians, Tunisian researchers and activists were prohibited from such information.[14]

Fourth, social media platforms employ discriminatory and unfair measures towards content from the Arab world.15 Facebook's organizational structure within the region speaks to systemic issues that reflect more broadly these orientalist tropes. While countries like Israel, and almost every European country has their own designated public policy head, the Middle East and North Africa region, despite the vast linguistic, state, religious, and cultural differences, are all lumped under one system of management. While Facebook maintains a broad "MENA" office in Dubai, they have a country specific office in Israel with their own public policy director (Jordana Cutler) who previously held the political positions of former adviser to Israeli Prime Minister Benjamin Netanyahu and Likud staffer. No such equivalent position exists for Palestinians or any other Arab country or diaspora.

Fifth, as most recently documented and recorded by 7amleh, users posting pro-Palestine content noted their audiences' views and reach was decreased. As Marwa Fatafta and Ahmed Shaheed and Benjamin Greenacre document in their contributions to this collection, these systemic policies have elevated and prioritised Israeli content and takedown requests.[16]

Platform moderation is a key area where these discriminatory politics play out. Jillian York's recent book *Silicon Values*[17] outlines the free speech implications of the choices made by these social media corporations. In one of her interviews with former platform content moderation experts, an anonymous former Facebook moderator said that when confronted about harms faced by groups in the region countries, Facebook would simply refuse to develop policies for those threats: "this kind of policy would never get any face time with the policy team because they were always busy with...whatever was prioritized by countries like Germany and the US."

When Orientalism goes online: Understanding *digital orientalism*

The recent evidence of censorship of pro-Palestine content in May 2021 was dismissed as merely "technical errors" by Facebook. This aligns with the more insidious patterns of systemic design discrimination, exacerbated by lack of resources and discriminatory policies. This has set the stage for situational crises in the Arab world, especially with the unfolding escalations of digital orientalism that pro-Palestine voices face in the midst of real-world repression.

While 2021 has been the year the concept of segregation and the systemic discrimination of Palestinians within Israel and the Palestinian territories has started to gain mainstream currency, Palestinian activists have begun to label injustices they claim they face online, both from social media companies and the Israeli Internet infrastructure that controls the flow of the Internet to Palestinians as "digital apartheid."[19] In this framework,

Table: Types of Platform Bias in the Arab World

Type	Definition	First Reported[18]	Platforms Involved
The removal of pro-democracy content	Relying on algorithmic or human moderation to remove content that does not go in line with platforms' "community standards."	2011	Facebook and YouTube
Restricting and deleting accounts of activists	Suspending accounts temporarily or deleting them permanently because they violated platforms' "community standards."	2011	Facebook
Limiting data access	Denying access to platform data despite providing them to Western researchers and civil society activists.	2019	Facebook
The lack of measures and resources to Arabic content	Not employing the same policies and measures that are applied in Western countries.	2021	All platforms
Reach reduction to activists' content	Adjusting the algorithms to reduce the reach of a certain type of content.	2021	Instagram and Facebook

Palestinian rights advocates argue that these forms of online discriminations are a continuation of the systemic forms of segregation, discrimination and abuses the Israeli authorities subjugate Palestinians to, only within the online realm. While we believe the term digital apartheid is correctly applied by Palestinian rights advocates within the broader struggles for rights and dignity against the Israeli state, we situate the problem more broadly into a more regional framework of digital orientalism. The discrimination and failures inherent in digital orientalism must be recognised as part of the growing strains of our online world. For too long this digital orientalism has been plaguing MENA countries, and in the case of Palestine, the platforms and their policies are in their ways contributing to events that the United Nations and leading human rights organisations say amount to war crimes.20

This is why we place the flaws in Facebook's policies into the context of more traditional media studies discourses. The arguments put forward by Edward Said in his last book within his *Orientalism* trilogies,[21] *Covering Islam,* broaches the question of the power of American and European media to shape perceptions of Islam and the countries of the region. Said's central argument was that media language builds and maintains stereotypes, and attempts to turn these western frameworks to describe a foreign culture into objective truths. We also see the parameters by which social media companies are policing speech and how this wields similar subjective power to build and maintain those same stereotypes. Said's focus on media coverage of the 1979 Revolution in Iran, for example, centred on the media's creations of notions of a "penchant for Shiite martyrdom", the "return of Islam" created as Western tropes that supplanted the causes of the Iranain movement and tried to speak for Iranians themselves. This language, Said argued, obscured the complexities and contradictions within the region and within Islam itself.

We see this echoed now through the new gatekeepers of information and culture: social media platforms. In today's new digital orientalism, the media narratives of Said's era have been adopted by the new gatekeepers of news and information and incorporated into the community guidelines of social media platforms. For example, Facebook's application of *Dangerous Individuals and Organisations* is a quagmire of problematic applications within the Middle East and North African region. The Western-centric origins of the policies that determine who these individuals and organisations mean that Facebook relies on the US Department of State's Foreign Terrorist Organisation's (FTO) list for its removal of accounts and content. The FTO list overwhelmingly includes a majority of Islamist terrorist entities,[22] as opposed to other lists like the United Nations Security Council terrorist list whose designations have a more global and religiously diverse distribution.[23]

This has become a major problem for freedom of expression, often hindering mere speech about events or news related to these designated FTO entities, who often are the topics of everyday life and governance in the region.[24] As Marwa Fatafta notes in this collection, Facebook has been known to equate words such as *Shaheed,* which is a common word within Islamic, Persian and Arabic lexicon as part of their *Dangerous Individuals and Organisations* policy.[25] Shaheed is a generic term for martyrs in Arabic, but the company's content moderation implementation automatically equates the word with terrorism, feeding into the Orientalist or Islamophobic conception of equating Islam with terrorism.

Digital Orientalism and Palestine

We now examine these issues in the context of the May 2021 crisis transpiring in Israel and Palestine in response to the movements against the forced evictions in the Palestinian Jerusalemite neighborhood of Sheikh Jarrah. Social media became central in two parts during the recent crisis: firstly, as protest and advocacy mechanism for Palestinian rights; and secondly, as a means of documentation of possible war crimes.

Since 2016, the digital repression against pro-Palestine content has been on the rise.[26] Mounting evidence throughout May 2021 has demonstrated a continued pattern of online discrimination by platforms,[27] one that

the Palestinian based digital rights organisation *7amleh* has been documenting for years.[28]

The Israeli Cyber Unit has indicated in the past that 85 percent of their government requests to "remove content deemed harmful or dangerous" from platforms such as Facebook, Google and Twitter are accepted.[29] During September 2016, Facebook complied with Israel's threats to block its platform in the country if Facebook did not comply with the deletion orders.[30] This move resulted in the deletion of many accounts of Palestinian activists and journalists.[31] Following online protests against Facebook, the company retrieved these accounts and apologised for this action.[32] This trend repeated itself in 2021. On 13 May 2021, the Israeli Justice Minister held a Zoom meeting with Facebook and TikTok executives to urge them to remove "anti-Israel" content.[33] The power of this Israeli pressure has been well documented and felt. Hundreds of accounts, pages, and groups associated with Palestinian activists and media outlets have been documented as deleted or their content removed.[34]

As mounting evidence shows the increasing erosion of rights for Palestinians, including the right to protest, the right to life, and worship, the coinciding online censorship has cemented fears that the systems of Israeli repression are being replicated online. While official statements by Facebook have resorted to explaining the initial issues as technical "glitches", digital rights activists have shared statements showing their dissatisfaction with Facebook's accountability and investigations.[35] Internal leaks have revealed there are more systemic issues at play to explain the censorship on Facebook's platform's, beyond mere technical errors. Hashtags such as *"Al Aqsa Mosque"* were systematically being blocked for reasons unrelated to the original technical glitch Facebook announced as causing the problems on Instagram.[36] Further investigations into Facebook reveal problematic and discriminatory policies are at play by Facebook. This is especially true in their development of policies surrounding speech critical or against "Zionism" which is the political ideology of Israel which in turn limits pro-Palestinian speech.[37]

Facebook announced that the ensuing situation that started in Sheikh Jarrah and led to the aerial bombardment in Gaza has led it to develop an Israel-Palestine crisis centre.[38] There is skepticism that the centre will not do much besides further embolden existing pro-Israeli and anti-free speech policies.[39] But there is hope that the recent uptake of the Arabic digital rights movement by civil society and media to pressure and seek accountability from these companies will lead to a shift in policies and prioritisation in the region. This public relations crisis for Facebook has highlighted that Arabic content moderation policies in the region validate a theory of digital orientalism. Users in this region are systemically subjected to a second-tier status as users with free speech and community support that is less than other regions and languages.

Conclusion: The Future of Arabic Content Moderation

When Arab activists noticed the systematic repression by social media platforms to suppress pro-Palestinian opinions, they took several steps to continue expressing their voices online. They promoted a campaign to downgrade the rating of Facebook on Google Play and Apple Store;[40] they used petitions, open letters, and articles to pressure social media companies to stop their algorithmic oppression;[41] they innovatively manipulated algorithms by tweaking the written Arabic text by either adding asterisk between letters, removing a letter from a word, adding "tanween" to words and hashtags, or changing the order of the letters.[42] One innovative approach has been to us the old dot-less Arabic cryptology tools. Social media AI is trained to read and analyse the standard Arabic letters — the one with dots in them — and this dot-less Arabic text tactic prevents the take downs of online content. While these measures seem promising, the digital repression facilitated by platforms' machine learning remains concerning. This chess-like game between Arab activists and platforms' architectural design is unbalanced and unfair. All the tactical innovations by Arab activists remain reactive and defensive against discriminatory systems that deprioritize their speech.

The digital orientalism of pro-Palestine content has led to silencing and censoring the voices of hundreds of thousands of the Internet's Arab users and their networks. They have also assisted Israel in erasing or drawing attention away from evidence of Israel's war crimes and human rights violations, and weakening the campaigns for Palestinian solidarity. Despite the grim events, the mobilization around Sheikh Jarrah has succeeded in generating an unprecedented amount of interest in the unfair practices and design of Arabic content moderation, both within social media discourses, organic campaigns to protest policies, and media coverage.

We contextualised the new digital orientalism of platform governance within the previous media theory framework of orientalism that Edward Said conceived. This theory of media orientalism has underpinned much of the Islamophobic media tropes that pervade Western society, where it has seeped into the policies of social media companies, the new gatekeepers of information. The same colonial infrastructures that subjugate and repress Palestinians in Israel in a apartheid state manifest themselves in the unequal conditions afforded to the users and preferences of Israel. As Zuckerberg alluded to in his 2012 pitch to investors, these voices cannot be ignored.

Endnotes

1 In a letter to investors during Facebook's Initial Public Offering (IPO)

2 Adrian Chen, "Mark Zuckerberg Takes Credit for Populist Revolutions Now That Facebook's Gone Public," Gawker, February 2, 2012, http://gawker.com/5881657/facebook-takes-credit-for-populist-revolutions-now-that-its-gone-public.

3 B'Tselem, "A Regime of Jewish Supremacy from the Jordan River to the Mediterranean Sea: This Is Apartheid," B'Tselem, January 12, 2021, https://www.btselem.org/publications/fulltext/202101_this_is_apartheid; Human Rights Watch, "Israeli Authorities and the Crimes of Apartheid and Persecution | HRW," April 27, 2021, https://www.hrw.org/report/2021/04/27/threshold-crossed/israeli-authorities-and-crimes-apartheid-and-persecution.

4 Hunt Allcott and Matthew Gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives 31, no. 2 (May 1, 2017): 211–36, https://doi.org/10.1257/jep.31.2.211.*

5 Paul Mozur, "A Genocide Incited on Facebook, With Posts From Myanmar's Military," New York Times, October 15, 2018, https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html.

6 Max Fisher, "Inside Facebook's Secret Rulebook for Global Political Speech - The New York Times," New York Times, December 27, 2018, https://www.nytimes.com/2018/12/27/world/facebook-moderators.html.

7 The page would later be credited with mobilising the revolutionary momentum that would remove Hosni Mubarak in 25 January 2011.

8 Anver Emon, Ellen Lust, and Audrey Macklin, "We Are All Khaled Said: An Interview with the Administrators of the Facebook Page That Fueled the Egyptian Revolution," *Boston Review* 3, no. 1–2 (2011), https://bostonreview.net/archives/BR36.6/khaled_said_facebook_egypt_revolution.php.

9 Jillian C. York, *Silicon Values : The Future of Free Speech Under Surveillance Capitalism* (Verso, 2021), 70.

10 William Lafi Youmans and Jillian C. York, "Social Media and the Activist Toolkit: User Agreements, Corporate Interests, and the Information Infrastructure of Modern Social Movements," *Journal of Communication* 62, no. 2 (April 1, 2012): 315–29, https://doi.org/10.1111/j.1460-2466.2012.01636.x.

11 Abdul Rahman Al Jaloud et al., "Caught in the Net: The Impact of 'Extremist' Speech Regulations on Human Rights Content," Electronic Frontier Foundation, May 30, 2019, https://www.eff.org/wp/caught-net-impact-extremist-speech-regulations-human-rights-content; Sarah El Deeb, "History of Syria's War at Risk as YouTube Reins in Content," *AP NEWS,* September 13, 2017, sec. AP Top News, https://apnews.com/d9f1c4f1bf20445ab06cbdff566a2b70; Kate O'Flaherty, "YouTube Keeps Deleting Evidence of Syrian Chemical Weapon Attacks," *Wired UK*, 2018, https://www.wired.co.uk/article/chemical-weapons-in-syria-youtube-algorithm-delete-video.

12 Mona Elswah and P. N. Howard, "The Challenges of Monitoring Social Media in the Arab World: The Case of the 2019 Tunisian Elections," Data Memo 2020.1, Computational Propaganda Research Project (Oxford Internet Institute, University of Oxford, 2020), https://comprop.oii.ox.ac.uk/research/posts/the-challenges-of-monitoring-social-media-in-the-arab-world-the-case-of-the-2019-tunisian-elections/.

13 The Facebook Ad library was released in 2019 to be a hub where Facebook can show its running ads. It is also used to archive political ads and present additional information on them (e.g., target audience, budget, sponsor, etc.). However, achieving political ads is not active in all countries and it is not enabled in the Arab region.

14 AccessNow, "Open Letter to Facebook on the Upcoming Tunisian Elections of 2019," *Access Now* (blog), September 2, 2019, https://www.accessnow.org/open-letter-to-facebook-regarding-the-upcoming-tunisian-elections-of-2019/.

15 Masaar, "Statement from Global Civil Society on the Impact of Facebook, Google and Twitter," *Massar* (blog), January 22, 2021, https://masaar.net/en/statement-from-global-civil-society-on-the-impact-of-facebook-google-and-twitter-concern-for-democracy-and-human-rights-must-not-end-at-the-uss-borders/.

16 Ryan Mac, "Instagram Labeled One Of Islam's Holiest Mosques A Terrorist Organization," BuzzFeed News, May 12, 2021, https://www.buzzfeednews.com/article/ryanmac/instagram-facebook-censored-al-aqsa-mosque.

17 York, *Silicon Values : The Future of Free Speech Under Surveillance Capitalism,* 20.

[18] This indicated to the first time — to our knowledge— this was reported by civil society organizations and media. That does not mean that these forms of biases were not taking place earlier.
[19] Palestine based digital rights activists at 7amleh have called this a "digital divide" in the past, but however, as of May 2021, 7amleh has been using the term "digital apartheid" (7amleh, 2017; Nashif, 2017).
[20] United Nations, "UN Human Rights Chief Appeals for De-Escalation in Israel-Palestine Crisis," UN News, May 15, 2021, https://news.un.org/en/story/2021/05/1092012.
[21] Edward W. Said, *Covering Islam: How the Media and the Experts Determine How We See the Rest of the World* (London: Vintage, 1997).
[22] US Department of States, "Foreign Terrorist Organisations," https://www.state.gov/foreign-terrorist-organizations/
[23] United Nations, "United Nations Security Council Consolidated List," https://scsanctions.un.org/consolidated/
[24] Lawrence Pintak, "The Trump Administration's Islamophobic Holy Grail," Foreign Policy, February 22, 2017; Isobel Cockerell, "Instagram Shuts down Iranian Accounts after Soleimani's Death," Coda Story, January 10, 2021, https://www.codastory.com/authoritarian-tech/instagram-iran-soleimani/; ARTICLE19, "Turkey: ARTICLE 19's Submission to the Facebook Oversight Board," *ARTICLE19, May 4, 2021.*
[25] Layla Mashkoor, "Sheikh Jarrah Content Takedowns Reveal Pattern of Online Restrictions in Palestine," The National News, May 10, 2021, https://www.thenationalnews.com/mena/sheikh-jarrah-content-takedowns-reveal-pattern-of-online-restrictions-in-palestine-1.1220037.
[26] Access Now, "Analysis: Facebook Zionism Hate Speech Policy Proposal," *Access Now* (blog), March 2, 2021, https://www.accessnow.org/facebook-hate-speech-policy-zionism/.
[27] 7amleh, "The Attacks on Palestinian Digital Rights: Progress Report," 7amleh, May 21, 2021.
[28] 7amleh, "Facebook and Palestinians: Biased or Neutral Content Moderation Policies?," October 29, 2018, https://7amleh.org/2018/10/29/7amleh-releases-policy-paper-facebook-and-palestinians-biased-or-neutral-content-moderation-policies.
[29] Shahar Ilan, "Israeli Official Reports Increased Cooperation on Removing Content from Social Media," CTech, December 29, 2017, https://www.calcalistech.com/ctech/articles/0,7340,L-3728439,00.html; Anan Abu Shanab, "Connection Interrupted: Israel's Control of the Palestinian ICT Infrastructure and Its Impact on Digital Rights," *7amleh,* January 31, 2019, 44.
[30] Glenn Greenwald, "Facebook Says It Is Deleting Accounts at the Direction of the U.S. and Israeli Governments," The Intercept, December 30, 2017, https://theintercept.com/2017/12/30/facebook-says-it-is-deleting-accounts-at-the-direction-of-the-u-s-and-israeli-governments/.
[31] Ylenia Gostoli, "Is Facebook Neutral on Palestine-Israel Conflict?," Al Jazeera, December 26, 2016, https://www.aljazeera.com/news/2016/09/26/is-facebook-neutral-on-palestine-israel-conflict/.
[32] Sophia Hyatt, "Facebook 'Blocks Accounts' of Palestinian Journalists," 2016, https://www.aljazeera.com/news/2016/9/25/facebook-blocks-accounts-of-palestinian-journalists.
[33] Emily Birnbaum, "Facebook Meets with Israeli and Palestinian Officials to Discuss Online Hate Speech, Threats as Violence Escalates," May 14, 2021, https://www.politico.com/news/2021/05/14/facebook-israel-palestine-hate-speech-488400.
[34] Access Now, "Analysis: Facebook Zionism hate speech policy proposal". *Access Now* (blog). March 2, 2021, https://www.accessnow.org/facebook-hate-speech-policy-zionism/.
[35] Matthew Ingram, "Social Networks Accused of Censoring Palestinian Content - Columbia Journalism Review," Columbia Journalism Review, May 19, 2021, https://www.cjr.org/the_media_today/social-networks-accused-of-censoring-palestinian-content.php.
[36] Ryan Mac. "Instagram Labeled One Of Islam's Holiest Mosques A Terrorist Organization." BuzzFeed News, May 12, 2021. https://www.buzzfeednews.com/article/ryanmac/instagram-facebook-censored-al-aqsa-mosque.
[37] Sam Biddle, "Facebook's Secret Rules About the Word 'Zionist' Impede Criticism of Israel," The Intercept, May 14, 2021, https://theintercept.com/2021/05/14/facebook-israel-zionist-moderation/.
[38] Elizabeth Culliford, "Facebook Deploys Special Team as Israel-Gaza Conflict Spreads across Social Media," Reuters, May 19, 2021, https://www.reuters.com/technology/facebook-running-special-center-respond-content-israeli-gaza-conflict-2021-05-19/.
[39] Marwa Fatafta, "Palestine-Israel Facebook Special Operations," Tweet, *Twitter,* May 20, 2021, https://twitter.com/marwasf/status/1395465697908727809.
[40] murdockism, "1 Star Facebook Ratings," Tweet, *Murdockism Tweet* (blog), May 18, 2021, https://twitter.com/murdockism/status/1394790980679782402.
[41] SMEX, "#SavePalestinianVoices: Tech Companies Must Stop Silencing Palestinian Content," Petition, Change.org, 2021, https://www.change.org/p/facebook-savepalestinianvoices-tech-companies-must-stop-silencing-palestinian-content.
[42] Mada Masr, "The Arab Revenge," Mada Masr, May 18, 2021, https://www.madamasr.com/ar/2021/05/18/feature/%d8%b3%d9%8a%d8%a7%d8%b3%d8%a9/%d8%a7%d9%84%d8%a7%d9%ae%d8%b3%d8%a7%da%ba-%d8%b5%d8%af-%d8%a7%d9%84%d8%a7%d9%84%d9%87-%d9%ae%d9%88%d8%b1%d9%87-%d8%a7%d9%84%d9%ae%d9%af%d8%a7%d8%b7-%d8%a7%d9%84%d8%b9%d8%b1%d9%ae%d9%89%d9%87/.

Official Foreign Influence Operations:
International Broadcasters in the Arab Online Sphere

Alexandra A. Siegel, University of Colorado - Boulder

International broadcasters, or state-funded media aimed at foreign publics, have long been an integral component of public diplomacy and foreign policy for authoritarian and democratic regimes alike. From Soviet use of Radio Moscow to spread communist ideology abroad beginning in the late 1920s to US sponsorship of Radio Free Europe and Radio Liberty during the Cold War, international broadcasting has often been deployed to shape global narratives and advance states' strategic goals.[1] In the aftermath of 9/11, international broadcasters—including the US-funded Alhurra and Radio Sawa—targeted Arabic speaking audiences to shape narratives in the MENA region. These outlets compete with regionally funded outlets such as Qatar's Al Jazeera and Saudi Arabia's Al-Arabiya, as well as national broadcasters.[2]

These state-funded media outlets have successfully adapted to the digital age, running influence campaigns both through traditional media channels and online. Such operations are clearly visible in the Arab online sphere, where international broadcasters have cultivated large audiences on Facebook, Twitter, YouTube, Instagram, and other platforms. State-sponsored media accounts have used targeted advertising on Facebook, Twitter, and Instagram, as well as tactics like clickbait headlines and hashtag campaigns, to gain followers and spread their narratives across platforms.[3]

From RT (formerly Russia Today)'s efforts to shape the regional narrative on the Syria conflict to Iran's Al-Alam campaigns to portray Iran as a dominant regional power, online campaigns by state media outlets are increasingly used to help foreign state actors advance their goals in the Arab World. While bots, trolls, and sock-puppets receive the lion's share of scholarly and journalistic attention—and are discussed in detail in contributions from Alexei Abrahams, Andrew Leber, Marc Owen Jones, Shelby Grossman, and Renee DiResta—international broadcasters are an important but understudied tool deployed in online influence operations. These accounts are extremely popular, often receiving some of the highest levels of engagement in the Arabic-language online sphere. Understanding the reach and influence of state-sponsored media accounts is perhaps particularly consequential in the Arab World, where trust in domestic media sources is relatively low.[4]

Labeling State-Backed Media

Recognizing the potential harms of foreign state-media operations in diverse global contexts, social media platforms have developed policies to label content produced by state-controlled media. YouTube was the first platform to label state-sponsored accounts in 2018, with Facebook, Twitter, and Instagram following in June, August, and September 2020 respectively. Facebook stated that this policy was enacted to "to provide an extra layer of protection against various types of foreign influence in the public debate."[5]

However, like all forms of content moderation, labeling state-sponsored media accounts is not a clear-cut task. As a result, labels have been inconsistently applied on several dimensions. First, only certain state-controlled media outlets have been labeled, and the list of flagged accounts differs across platforms. For example, Iran's English language PressTV accounts are labeled by Facebook and YouTube, but not by Twitter. Both YouTube and Facebook have periodically blocked the account entirely for violating its terms of service, most recently in late March 2021.[6] Other popular state-controlled outlets, including Iranian, Turkish, and Israeli international broadcasters have not been labeled at all. Moreover, Western international broadcasters appear to be exempt from labeling, presumably because they have "sufficient editorial independence,"[7] though how this is determined remains unclear. Additionally, labels are applied differently across types of content, often not appearing in platform search results or on live content or "stories."[8]

Here I examine four international broadcasters that are particularly popular in the Arab online sphere: Russia's RT Arabic, China's CGTN Arabic, Iran's Al-Alam, and Turkey's TRT Arabi. Both RT and CGTN have been labeled as state controlled foreign media by social media platforms, while Al-Alam and TRT have not. Analyzing about 700K tweets and 500K public Facebook posts produced by these international broadcasters' accounts, I show a decrease in both followers and engagement in the aftermath of the platforms' labeling policy for RT and CGTN Arabic, relative to the unlabeled Al-Alam and TRT accounts. Before presenting this descriptive analysis, I first provide a brief overview of each outlet's origin and presence across Arabic-language social media platforms.

RT (formerly Russia Today) launched in 2005 with the stated goal of bringing "the Russian view on global news," and launched its Arabic language channel in 2009. But RT soon changed its slogan to "Question More" on both its English and Arabic language channels, framing itself as an alternative to "biased" Western news sources. RT's Arabic language website does not include a mission statement and simply lists it as a subsidiary of RIANovosti while highlighting the satellite stations through which it transmits its broadcasts to the region.[9] Vladimir Putin has described RT's purpose as "break[ing] the Anglo-Saxon monopoly on the global media."[10] RT's editor in chief has even stated that RT is "conducting [an] information war," playing a role as vital as the Ministry of Defense. She elaborated that the outlet's strategy is to cultivate an audience that considers RT a source for trusted news, with the goal of helping the Russian state disseminate its message during critical moments.[11] RT Arabic has 17.5 million followers on Facebook and 5.2 million followers on Twitter. Recent research suggests that RT Arabic has been particularly influential in spreading online narratives on the Syria conflict, surpassing engagement of mainstream news outlets in both Arabic and English.[12] These social media posts portray Russia as effective in fighting extremism, accuse the US of committing human rights abuses, and highlight Western military failures.

China's CGTN (formerly CCTV) launched its free-to-air Arabic-language international channel in 2009, announcing that the channel would "serve as an important bridge to strengthen communication and understanding between China and Arab countries."[13] The outlet soon created accounts on Twitter and Facebook, YouTube, and Instagram, which are blocked inside China. The CGTN Arabic Twitter account has about 700K followers, while its Facebook page has over 15 million followers. CGTN Arabic accounts often produce anti-Western content that advances China's foreign policy interests. Following the outbreak of COVID-19 as former President Trump blamed China for the spread of the virus, CGTN Arabic began to push content emphasizing that the pandemic started in the US, criticizing the US pandemic response, and charging the US with human rights abuses.[14]

Iran's Al-Alam is an Arabic-language channel, which relies on financial and logistic support from the Islamic Revolutionary Guard Corps.[15] In February 2003, Al-Alam began broadcasting from Iran into Iraq, with the goal of targeting Iraq's majority Shia population.[16] The following year, Al-Alam launched a website, ww.alalam.ir, and expanded its audience to target Shia Arabs more broadly. Al-Alam launched its public Facebook page in 2010, which now has about 6 million followers,[17] and has an active presence on Twitter, YouTube, and Instagram. The channel's goal, whether over the airwaves or online, is to diminish the influence of Iran's rivals in the region, while advancing Iran's foreign policy objectives. The outlet regularly highlights Iran's accomplishments and successes from technological advances to soccer victories. It also emphasizes pan-Islamic identity, downplaying sectarian and national identities and portraying Iran as the true defender of Islam. The channel frequently portrays Western countries as threats to Islam in the context of ongoing regional conflicts.[18]

Lastly, Turkey's TRT Arabic language channel was launched in 2010 to reach the Arabic-speaking and Islamic world with its broadcasts. The goal of the media outlet was to advance the AKP's transnational agenda "to exert a form of soft power in the MENA region."[19] Along these lines, Erdoğan announced at the launch of TRT's Arabic channel, that Turks and Arabs "share the same history, culture and civilization... They are like the fingers of a hand. They are as close as the

flesh and the nail of a finger."[20] The AKP government has used TRT to enhance its political and economic standing by "strategically constructing an attractive neo-Ottoman nation brand."[21] Alongside its satellite channel, TRT Arabi has also become popular on social media, with about 3 million followers on Facebook and 1.1 million followers on Twitter. These accounts disseminate a range of TRT Arabic media content highlighting Turkish strength and supporting Turkish foreign policy goals in the Arab online sphere.

Did labeling matter? The data.

Did platform labels impact the reach of state-sponsored accounts in the Arab online sphere? Examining account followers and engagement over time offers suggestive evidence that platform labels reduced the reach of accounts on Twitter and Facebook, relative to unlabeled accounts. To look at changes in follower counts over time on Twitter, I used archive.org snapshots of international broadcaster Twitter accounts' historical profiles.[22] As Figure 1 displays, while RT Arabic and CGTN Arabic's follower counts were growing consistently in the leadup to Twitter labeling the accounts in August 2020 (marked by a black vertical line in the plots below), the growth flattened and slightly decreased following the labeling. Each vertical blue line in the plot represents a snapshot of the account's follower count captured by the internet archive. This contrasts with the growth patters we see for Al-Alam and TRT Arabi, which were not labeled by Twitter, and experienced increased growth over the entire period.

Figure 1: Change in Followers Over Time Labeled State Media Accounts

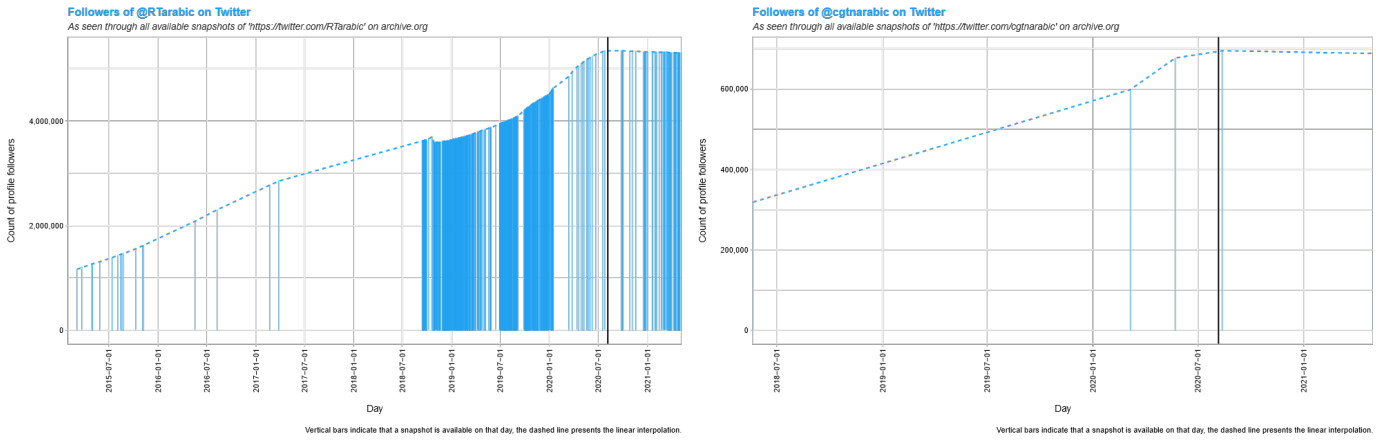
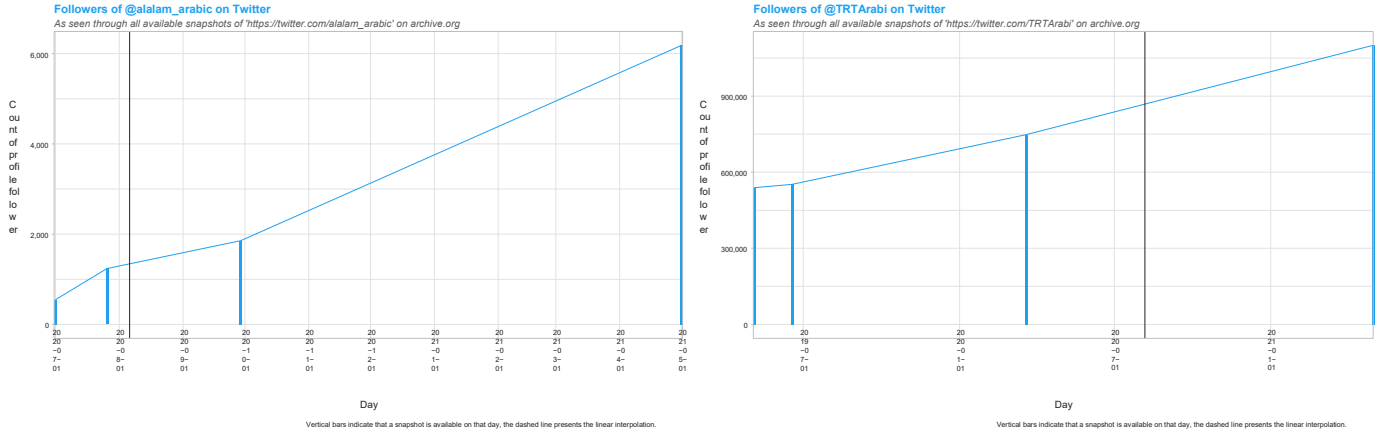


Figure 2: Change in Followers Over Time Unlabeled State Media Accounts



Using Facebook monthly follower counts at the time of each post obtained using the CrowdTangle API,[23] we see largely similar patterns. The growth in followers of the RT Arabic and CGTN Arabic public pages displayed in Figure 3 starts to level off in the aftermath of Facebook's labeling policy in June 2020, though the change is less immediate than what we observe in the Twitter data. Looking at unlabeled pages in Figure 4, we see that TRT Arabic's follower count continued to grow steadily, while Al-Alam's page gained followers sharply and then has a subsequent decline in followers. In these figures the announcement of Facebook's labeling policy is marked by a red vertical line.

Figure 3: Change in Followers Over Time Labeled State Media Pages

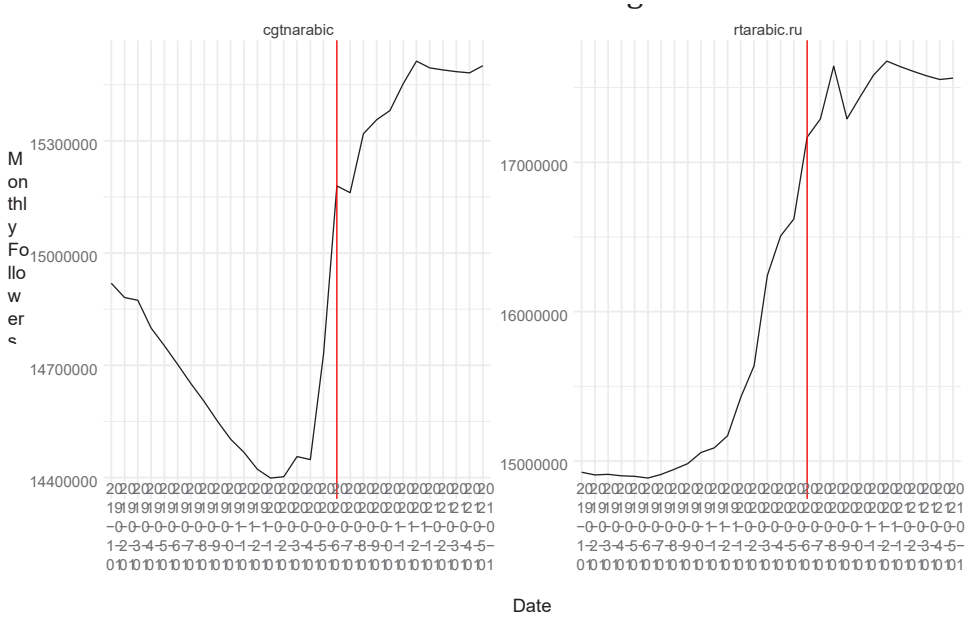


Figure 4: Change in Followers Over Time Unlabeled State Media Pages

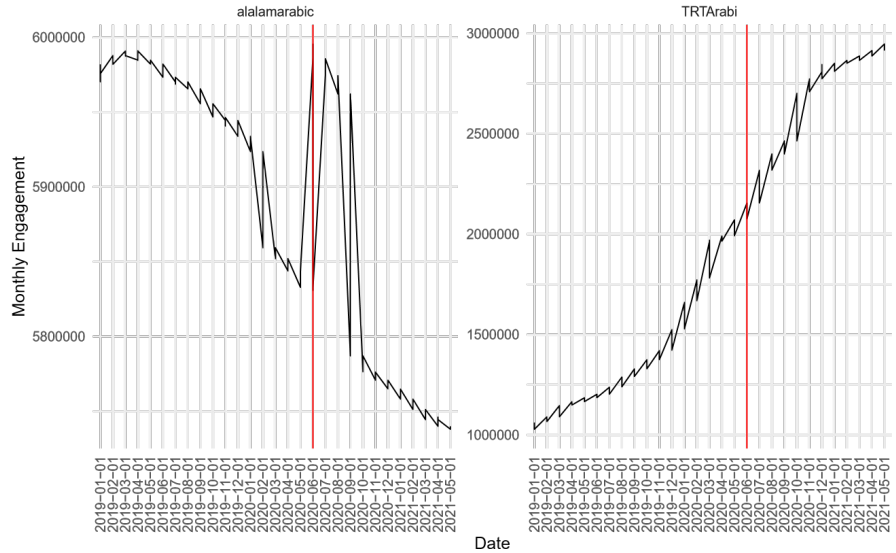


Figure 5: Change in Engagement Over Time Labeled State Media Accounts

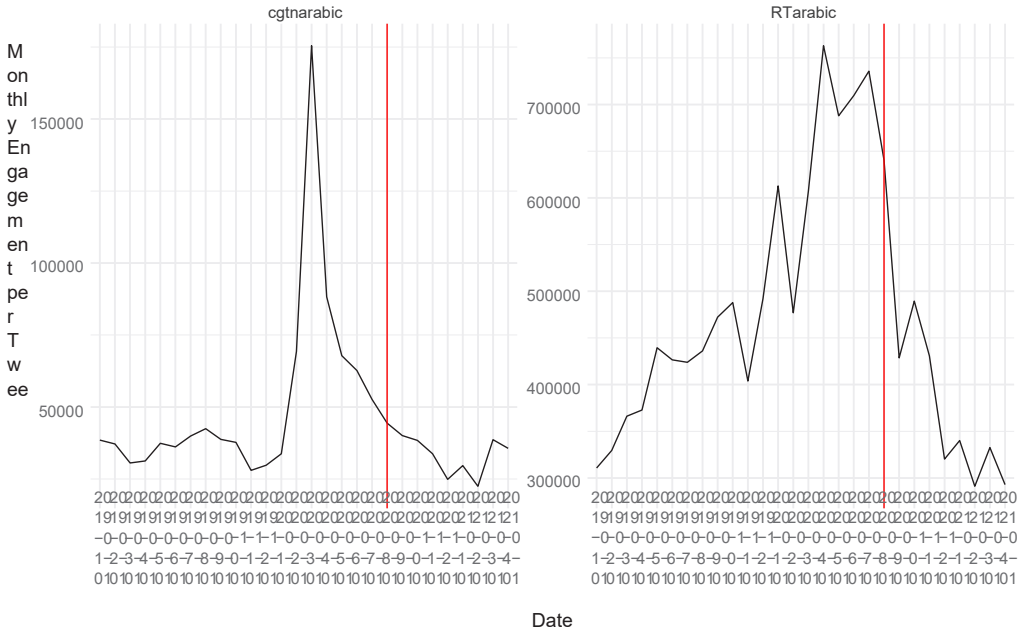
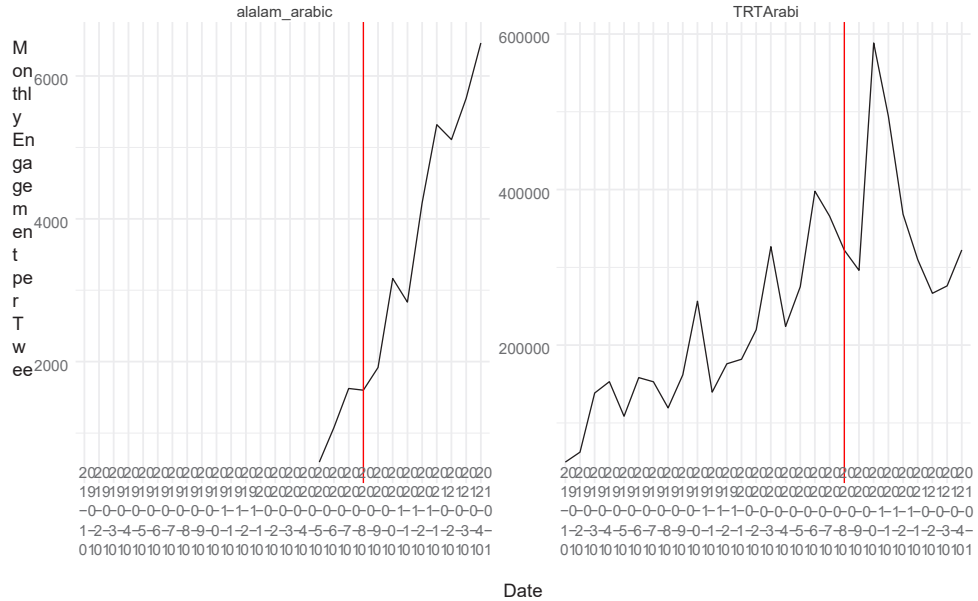


Figure 6: Change in Engagement Over Time Unlabeled State Media Accounts



Examining changes in engagement on Twitter over time, measured as retweets, likes, quote tweets, and comments collected with the academic Twitter API,[24] Figures 5 and 6 show a decline in engagement with RT Arabic immediately following the account labeling in August 2020 and a decline in engagement with CGTN Arabic, which begins before the account labeling and continues in its aftermath. By contrast, engagement with Al-Alam tweets continued to grow over the entire period, and engagement with TRT Arabi continued to grow and then declined in early 2021.

Changes in engagement with Facebook pages are less clear, with expected declines in engagement with CGTN Arabic and RT in the aftermath of Facebook's labels announcement, but also a similar decline in engagement with Al-Alam Arabic, which was not—to my knowledge—labeled by Facebook in this period. It is possible, however, that Facebook's restrictions of advertisements from state-sponsored outlets may have affected some of these accounts as well. This dramatic spike in activity followed by a decline could also be related to the use of inauthentic accounts to boost follower numbers.

Figure 7: Change in Engagement Over Time Labeled State Media Pages

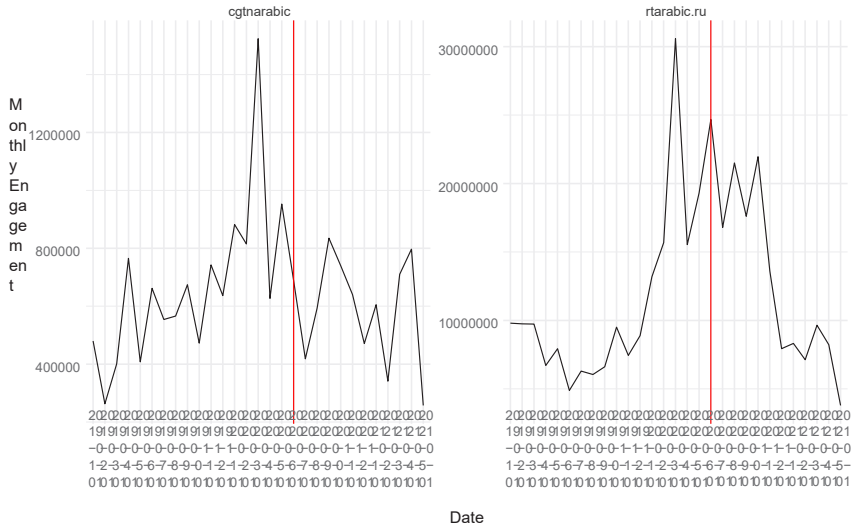
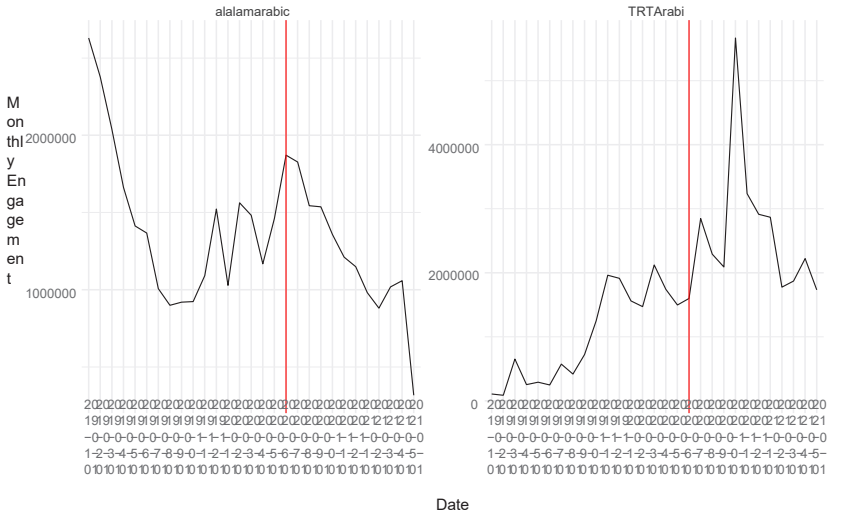


Figure 8: Change in Engagement Over Time Unlabeled State Media Pages



Together, publicly available social media data suggests that platform applications of labels to state-sponsored media accounts may have reduced follower counts and engagement in the Arab online sphere, relative to unlabeled accounts. This was particularly true on Twitter, where changes in follower counts and engagement dropped most dramatically for labeled accounts compared to unlabeled ones.

Discussion and Implications

Given the potential of social media platforms' policies to shape the visibility of state-sponsored content, this preliminary analysis raises important questions regarding how these policies are applied. Why are some state-sponsored accounts flagged while other similar platforms are not? In addition to decreasing engagement and follower counts, do platform labels make these sources less credible to their audiences? What proportion of these accounts' followers are authentic? Considering the large followings of international broadcaster accounts across social media platforms, future research on foreign influence operations should examine how these overt campaigns interact with more covert strategies—such as those explored in the contributions from Akin Unver, Mark Owen Jones, Shelby Grossman and Renee DiResta—as well as their impact on attitudes and behaviors in diverse contexts.

Endnotes

[1] Gary D. Rawnsley "To know us is to love us: Public diplomacy and international broadcasting in contemporary Russia and China." *Politics* 35, no. 3-4 (2015): 273-286.
[2] Deena Dajani, Marie Gillespie, and Rhys Crilley. "Differentiated visibilities: RT Arabic's narration of Russia's role in the Syrian war." *Media, War & Conflict* (2019)
[3] Megan M. Metzger, and Alexandra A. Siegel. "When State-Sponsored Media Goes Viral: Russia's Use of RT to Shape Global Discourse on Syria." *Working paper* (2021); Cook, Sarah. "Beijing's Global Megaphone." *Freedom House* (2020).
[4] Nadine Sika. "Contentious activism and political trust in non-democratic regimes: evidence from the MENA." *Democratization* 27, no. 8 (2020): 1515-1532.
[5] France 24, "Facebook Labels State-Controlled Media Posts, Will Block Ads," (June 4, 2020), https://www.france24.com/en/20200604-facebook-labels-state-controlled-media-posts-will-block-ads
[6] Press TV's Twitter account, accessed May 23, 2021: https://twitter.com/presstv/status/1377304326868197389?lang=en
[7] Facebook's content announcement, accessed May 23, 2021: https://about.fb.com/news/2020/06/labeling-state-controlled-media/
[8] Morgan Wack. "Inconsistencies in State-Controlled Media Labeling." Election Integrity Partnership. Election Integrity Partnership, October 6, 2020. https://www.eipartnership.net/policy-analysis/inconsistencies-in-state-controlled-media-labeling.
[9] Deena Dajani, Marie Gillespie, and Rhys Crilley. "Differentiated visibilities: RT Arabic's narration of Russia's role in the Syrian war." *Media, War & Conflict* (2019)
[10] Max Fisher. "In Case You Weren't Clear on Russia Today's Relationship to Moscow, Putin Clears It Up." *The Washington Post*. April 29, 2019. https://www.washingtonpost.com/news/worldviews/wp/2013/06/13/in-case-you-werent-clear-on-russia-todays-relationship-to-moscow-putin-clears-it-up/.
[11] "Question That: RT's Military Mission." Medium. DFRLab, November 23, 2018. https://medium.com/dfrlab/question-that-rts-military-mission-4c4bd9f72c88.
[12] Megan M. Metzger, and Alexandra A. Siegel. "When State-Sponsored Media Goes Viral: Russia's Use of RT to Shape Global Discourse on Syria." *Working paper* (2021)
[13] Sarah Cook. "Beijing's Global Megaphone." *Freedom House* (2020).
[14] Sarah Cook. "Beijing's Global Megaphone." *Freedom House* (2020).
[15] Massoumeh Torfeh. *The Role of Iran's Regional Media in its Soft War Policy*. Al Jazeera Centre for Studies, 2016.
[16] Khalaf M. Tahat, and Gilbert Fowler. "Iranian Propaganda in the Middle East: Al Alam" The World" as Model." *Southwestern Mass Communication Journal* 26, no. 2 (2011).
[17] Al-Alam's Facebook page. Accessed: May 23, 2021 https://www.facebook.com/alalamarabic/
[18] Khalaf M. Tahat, and Gilbert Fowler. "Iranian Propaganda in the Middle East: Al Alam" The World" as Model." *Southwestern Mass Communication Journal* 26, no. 2 (2011).
[19] Gökçen Karanfil. "Continuities and Changes in the Transnational Broadcasts of TRT." In *Television in Turkey*, pp. 151-171. Palgrave Macmillan, Cham, 2020.
[20] Omar Al-Ghazzi, and Marwan M. Kraidy. "Turkey, the Middle East & the Media| neo-ottoman cool 2: Turkish nation branding and Arabic-language transnational broadcasting." *International Journal of Communication* 7 (2013): 20.
[21] Omar Al-Ghazzi, and Marwan M. Kraidy. "Turkey, the Middle East & the Media| neo-ottoman cool 2: Turkish nation branding and Arabic-language transnational broadcasting." *International Journal of Communication* 7 (2013): 20.
[22] Code repository for scraping past follower counts using the internet archive. Accessed: May 23, 2021. https://github.com/ChRauh/PastTwitter.
[23] The Crowdtangle API home page. Accessed: May 23, 2021. https://www.crowdtangle.com/.
[24] Code repository for accessing the Academic Twitter API. Accessed: May 23, 2021. https://github.com/cjbarrie/academictwitteR,

Russian Digital Influence Operations in Turkey 2015-2020

Akin Unver, Ozyegin University and Ahmet Kurnaz, Çanakkale Onsekiz Mart University

The literature on online disinformation studies focuses disproportionately on the United States - especially on the 2016 Presidential elections – and has failed to generate an equally robust and diverse research agenda elsewhere.[1] Empirical studies have drawn on a very narrow pool of cases, with the overwhelming majority of the scientific and policy focus on what Russia is doing in the United States, or a handful of Western nations.[2] This impairs construction of a truly comparative and generalizable scientific inquiry, especially in terms of what disinformation (deliberate use of false information to deceive) or influence operations (deploying a mix of accurate, semi-accurate and false information to achieve strategic goals) mean for the broader world and international competition dynamics. To that end, the study of both fields is in need of longitudinal and comparative works: to provide perspective on how disinformation dynamics observed at one time are different than those at others; how dynamics observed in one country differ from those in other countries; and how operations conducted by different external actors vary. What's more, availability bias afflicts the wider disinformation studies field, as very few studies deal with the question of what the existence of disinformation means in relation to the cases where information manipulation doesn't exist. In this essay, we examine Russian information operations in Turkey as a first step towards addressing these shortcomings in the literature.

Turkey as a Case Study

Why Turkey? 'Buffer countries' or 'insulators' as defined in Regional Security Complex Theory (RSCT) are well-suited for such comparative work.[3] Although there are clear theoretical and methodological differences between how these two terms are studied, they both indicate countries that lie at the intersection of two or more large security communities. Such countries are usually not powerful enough to dominate either community, but also not weak enough to be dominated by either. To that end such countries are regularly influenced by multiple security communities, and their domestic power dynamics acutely reflect external security-related influences; in turn, these internal dynamics have significant impact on policy towards external security communities.

Turkey is one of those buffer or insulator countries.[4] Its imperial and Republican foreign policy were both heavily influenced by hedging and balancing dynamics against the Russian Empire, and then the USSR. Even as a NATO ally, Turkey competed with other NATO countries (most specifically Greece) and cooperated with the USSR (especially in building the Turkey's heavy industries in the 1970s) as circumstances dictated.[5] Although the end of the Cold War and the next two decades enabled Turkish policymakers to build a new security identity against a weaker Russia, the rise of an emboldened and revisionist Russian foreign policy after 2010 brought back structural balancing considerations for Ankara. Especially after the annexation of Crimea, the Russian military encroaching into the Black Sea, Syria, eastern Mediterranean and the Caucasus, Turkish decision makers increasingly found themselves dealing with situations that amounted to being strategically surrounded by Russia.[6] Ankara felt that NATO continually failed to provide sufficient security commitments against Russian encroachments, resulting in Turkish hedging and then bandwagoning with Russia.

Turkish foreign policy after 2014 can thus be described as multilateral hedging at a time of significant changes in the balance of power in its immediate environment (Russia-related), and also at the global level (China-related).[7] In addition, a growing strategic divide between the US and Europe (as well as within the EU itself) placed Turkey at the intersection of multiple strategic influences originating from Washington, London, Brussels, Berlin, Moscow and Beijing. Turkey was also embroiled in regional competition with countries such as Egypt, Saudi Arabia and the United Arab Emirates which actively engaged

in digital disinformation and influence operations. As a result, Turkey became a battleground for foreign influence operations, not just limited to Russia. Therefore, Turkey after 2014 is one of the most interesting case studies for the study of multilateral digital influence operations in general, and disinformation in particular.

Russian Digital Influence Events in Turkey: The Data

In a recently concluded project, our lab has focused on building a 'Russian influence event dataset' (RUSDAT) that collects social media data on such activities since 2014.[8] This paper updates the original 2019 publication, with new data which we continued to collect as Russian disinformation activities continued.

RUSDAT was built on several criteria. First, we focused on bilateral geopolitical events between Russia and Turkey, constructed a keyword corpus that contained terms and word combinations related to each high-profile strategic event and extracted all Twitter data that corresponded to those events. We then sorted them according to the amount of clean data we had after weeding out irrelevant posts (including tweets from brands, football clubs or Korean pop bands, which surprisingly often post local hashtags to rise into the trending topic list!). Finally, we ranked these events based on how much clean data we had on them and discarded cases that contained too few tweets (below 2 million) or had too much dirty data as percentage of the whole dataset. Ultimately, we focused on four of the most important events that also contained the highest volume and percentage of clean data to explore deeper, although initially discarded cases were retained within RUSDAT.

The 'clean' cases picked for study were Turkey's downing of a Russian jet in November 2015, Turkey's failed coup attempt in July 2016, the assassination of the Russian ambassador in Ankara in December 2016, and the S-400 negotiations between Turkey and Russia, which itself was separated into six key benchmarks (declaration of interest, declaration of no-cancellation, signing of the purchase memorandum, signing of the commercial agreement and final acquisition). Social media dynamics throughout the S-400 negotiations were particularly useful as each benchmark gave a clear idea on how influence operations and media response against them changed over time. Out of these cases, we were able to identify a distinct 'pro-Russian' influence cluster that encompassed a large network of Turkish-language real and sock puppet accounts, occasionally supporting the narratives of pro-Russian Turkish-language outlets Sputnik News Turkish, Aydınlık newspaper and Russia Today's Turkish-language news section.

The 2015 Downing of a Russian Jet: Distraction

Over the last few years, foreign observers of Turkey frequently asked, 'who lost Turkey?', meaning whose fault it was that Turkey had become so detached from the West.[9] Answers ranged from general NATO apathy towards Turkey's changing security environment after Crimea annexation, to European analysts blaming Trump, or American analysts blaming European resistance to Turkey's EU membership. This question can be better asked temporally: *when was Turkey 'lost'?* From a digital communication point of view, our study can pinpoint a single event: the Turkish downing of a Russian jet in November 2015.

Soon after the Russian jet was shot down, we began observing the emergence of two discursive clusters (or narratives). The Turkish version argued that the decision was justified because the unidentified jet had strayed too much into Turkish territory. A second cluster of tweets asserted that the jet was shot outside the Turkish airspace, and was thus, unjustified. As internal military investigations of both sides began yielding results that supported the first claim after a week, Russian outlets adopted an organized distraction tactic which originated in accounts associated with the Ministry of Defense (based on how the initial MoD tweets were spread across both English- and Turkish-language Twitter ecosystem), which focused on Turkey's alleged oil smuggling deal with ISIS.[10]

This distraction tactic, originally crafted in and disseminated by the Kremlin (based on its first appearance and subsequent diffusion patterns on Twitter), soon got picked up by international news and media agencies,

including those of other NATO countries. Successfully distracting the discussion away from the SU-24 incident, this became one of Kremlin's most efficient influence operations across the entirety of NATO countries, managing not only to divide and nullify NATO's countermeasures against Russian violations of NATO airspace, but also created a very significant wedge between Turkey and its Western allies, isolating Turkey in the short- to medium-term. Although both the Pentagon and the State Department had rejected Russian allegations of Turkey's ISIS-related oil smuggling,[11] the story was disseminated far and wide in Western capitals, ending conclusively only after Presidents Putin and Erdoğan met in August 2016. This meeting, where Turkey conceded defeat in its information war with Russia, was a major turning point in Turkey's relations with Russia.[12] The 'ISIS oil' story then disappeared entirely and immediately on Russian and Turkish-language Twitter.

The 2016 Failed Coup: Amplification

During Turkey's failed July 2016 coup attempt, Russian influence operations benefited significantly from the pre-existing and growing Turkish domestic skepticism towards the US and NATO. Some of this public skepticism was a result of growing strategic disagreements in Syria. The Turkish government's vocal complaint that neither the US nor Europe (with the exception of the UK) condemned the coup attempt during its early hours had a major rallying effect around the narrative that the coup was instigated by NATO.[13]

Throughout the coup attempt, all of the widest-spread disinformation instances had a domestic origin. But pro-Russian accounts did try to amplify the prevalent public sentiment that the coup was planned and orchestrated by pro-NATO cells within the military. That said, compared to other major instances, the activity of pro-Russian accounts throughout the failed military coup make up only a very small fraction (less than 1%) of the total engagement clusters observed. Even months after the failed coup attempt, pro-Russian accounts continued to sustain the narrative that it was NATO-affiliated groups that were behind the coup itself.

Assassination of Russian Ambassador: Silence

Five months later, Russian Ambassador Andrey Karlov was assassinated in Ankara, straining already fragile Turkish-Russian relations even further. Yet the pro-Russian influence ecosystem went completely dark, suggesting a centrally-planned full silence. Why? First, after the August 2016 Putin-Erdoğan meeting, the two countries had charted a common course to deconflict bilateral relations and Russia had no further interest in destabilizing Turkey. Second, the Russian government was already in close communication with Turkey to contain the damage of this incident as quickly as possible.[14] The entirety of the social media war that followed the Karlov assassination was domestic to Turkey, with arguments taking place between two pro-government clusters: one that viewed the assassination as 'justified' in the face of growing Russian attacks against pro-Turkish rebel groups in Syria, and the other, which advocated for calm and reconciliation in line with the mainstream government view. The assassination debate disappeared to a great extent on social media after only four days, suggesting a direct gag order by both Ankara and Moscow.

The S-400 Negotiations: Sustained Influence Operations

Finally, we explored the S-400 negotiations, as divided into six benchmarks: 10 October 2016 when Turkey and Russia declared that serious Presidential-level negotiations were underway over S-400 sales, Erdoğan's 10 March 2017 visit to Moscow to assert Turkey's commitment to S-400, 29 December 2017 commercial agreement between the two sides, 3 April 2018 President Erdoğan's statement on Turkey's 'point of no return' on S-400 purchase, 19 August 2018 President Putin's statement that deliveries could be made a year earlier than planned, and July 2019 when the first shipment of S-400 ground systems arrive in Turkey. The S-400 case demonstrates the explanatory power of measuring influence operations across a longer timeframe (in this case, across almost 3 years), as the Turkish media ecosystem changed to an important degree throughout this episode, as did many other national and international variables of interest.

Figure 1 - Longitudinal sentiment scores of positive and negative sentiment clusters; October 2016 – July 2019

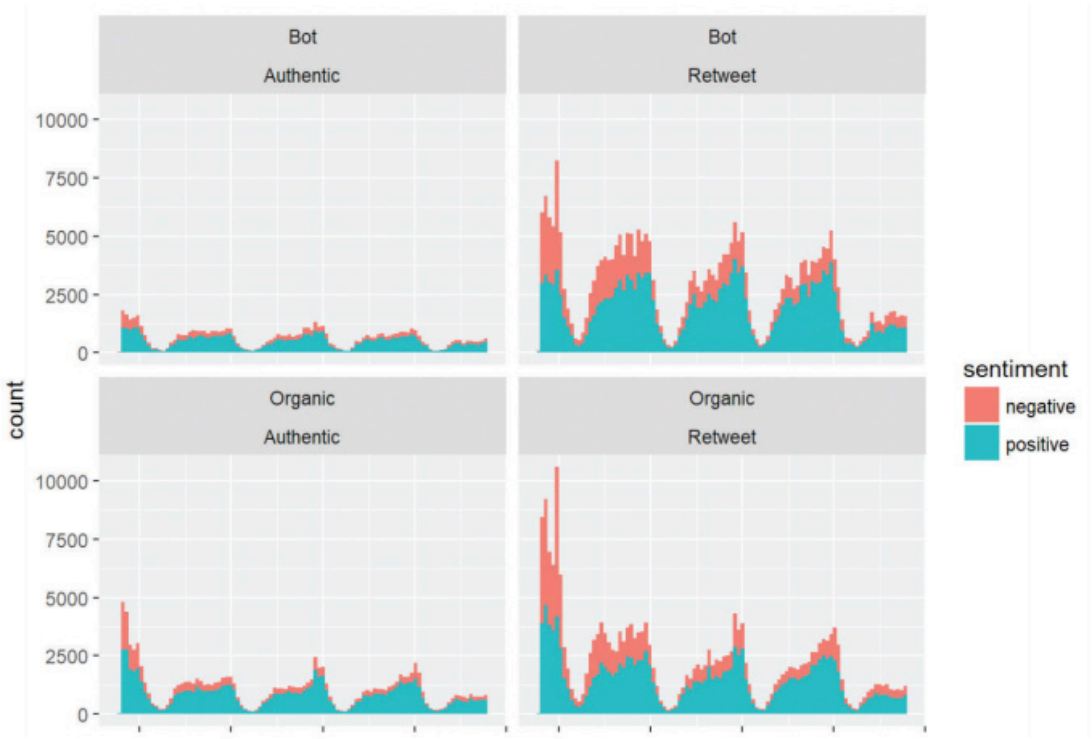


Table 1 - News outlets that form up the core of positive and negative sentiment clusters

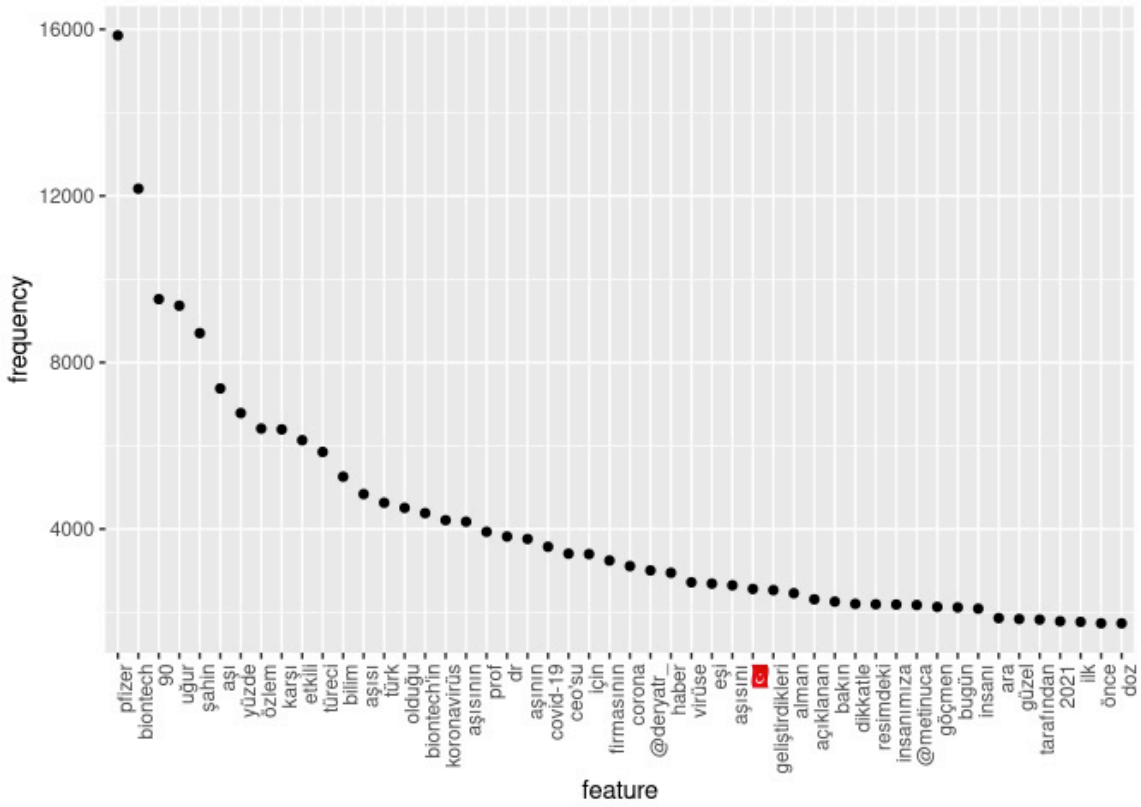
Negative Sentiment Domain	% Share of Aggregate Tweets	Positive Sentiment Domain	% Share of Aggregate Tweets
Hurriyet.com.tr	22.382	Sabah.com.tr	27.294
CnnTurk.com.tr	19.504	Yenisafak.com	21.953
Sozcu.com.tr	17.828	HaberTurk.com.tr	19.144
Cumhuriyet.com.tr	15.382	Ahaber.com.tr	13.048
T24.com.tr	13.449	Star.com.tr	8.032
bbc.com (Turkish)	6.486	OdaTV.com	5.392
DW.com (Turkish)	3.291	Tr.Sputniknews.com	5.019

The most important finding is the gradual transition of the Turkish-language sentiment scores (measured by deploying 'BERT Sentiment Analysis Turkish'[15]) associated with S-400s, from mixed (equal measures of positive and negative), to mostly positive across the six benchmarks we observe. This means that the overall outlook of the Turkish social media ecosystem towards the S-400s started off as skeptical and divided, and gradually became very positive towards these systems. The main words associated with the skeptical topic clusters reflect worries about interoperability of the Russian systems with NATO infrastructure, Turkey's existing NATO commitments and what the S-400 acquisitions would mean for Turkey's other major partnership in the F35 fighter jet program. In contrast, word clusters associated with the 'pro-S-400' sentiments reflect the importance of strategic autonomy, NATO's broader relevance for Turkey, and technical details that reflect the view that S-400s are 'better' anti-air systems than the Patriots. Over time, 'pro-S-400' topic clusters dominate the Turkish-language discussion with heavy involvement of pro-Russian and also pro-government accounts in Turkey.

Dynamics After 2019: COVID, Nagorno-Karabakh and Biden

Three main additional events triggered pro-Russian influence operations after 2019. The first was the emergence of COVID-19 and the onset of the global race for vaccines. The second major event was the 2020 Nagorno-Karabakh war between Azerbaijan and Armenia, in which Turkey and Russia served as major external stakeholders. The third major event was the election of Joe Biden and the proliferation of skeptical news reports from the pro-Russian ecosystem on his capacity to lead, or whether his election would really make a difference.

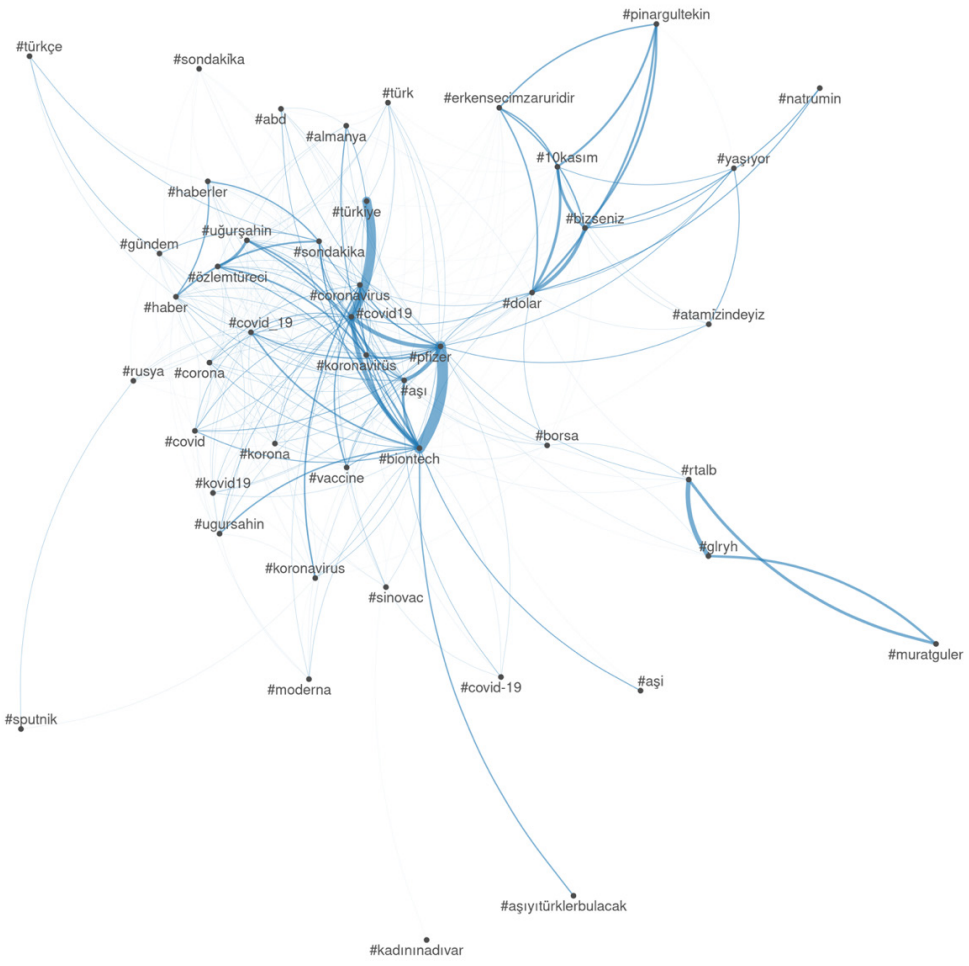
Figure 2 - Frequency hierarchy of the most popular features (terms) on Coronavirus



Predictably, the pro-Russian network in Turkish-language social media platforms (Twitter, Facebook and Instagram) sowed widespread confusion about the efficacy and side-effects of American and European vaccines, while remaining silent about the Chinese SINOVAC, and continually advertising the Russian Sputnik-V vaccine. This network published vaccine-related information which focused on vaccine skepticism, emphasizing the importance of getting vaccinated, although with a twist that always ends with a positive note about the affordability, availability and the efficacy of Sputnik-V. Further word clusters along this line focus on the positive international reception of the Sputnik vaccine and the cases of patients that were saved thanks to getting vaccinated. This information pushed back on BBC Turkish, Deutsche Welle Turkish and Fox Turkey articles that disseminated skeptical views of Sputnik V or the Sinovac/ Coronavac vaccines.

On Nagorno-Karabakh, Russian influence operations were extra careful, as the long-frozen conflict had been one of the core national interests of Turkey and had multi-partisan support among the Turkish voters. Since there was no domestic audience to which Russia could play on this matter, most pro-Russian accounts focused instead on the need to stabilize the Karabakh region and

Figure 3 - Feature frequency network of Coronavirus-related terms



posted themes related to the 'possibility' of an Armenian-Azerbaijan reconciliation with the joint oversight of Turkey and Russia. During the conflict itself, however, these accounts pursued a distinct pro-Armenian line, occasionally sharing low-diffusion disinformation content about the course of the conflict, including false accounts of attacks, casualties and clashes.

Finally, the entirety of the pro-Russian information ecosystem turned uncharacteristically over-active after the election of Biden, regularly disseminating fake news about his physical and mental fitness and questioning his ability to lead. Additionally, this ecosystem had been using key events, such as Biden's recognition of the Armenian genocide, to draw a wedge between him and President Erdoğan. In one such instance, a large bot campaign pushed the argument that the Turkish government must file an international lawsuit against Biden on the grounds of 'hate speech', in exchange for Biden's recognition of the genocide. Another major campaign focused on pinpointing Biden as the main culprit behind the CAATSA sanctions issued against Turkey by the US Congress under the Trump administration.

Conclusion and Implications

To sum up, this project has so far yielded nuanced results that show a more cautious, more context-specific and more 'under-the-radar' digital influence strategy on the part of Russia. We hypothesize that Russia's relative caution in influencing Turkish digital media ecosystem owes to the fact that Turkey, as an insulator country, is indeed divided between multiple foreign influence strategies and possible Russian interpretation that further destabilization of this ecosystem would trigger a backlash. This is interesting because the majority of the US-centric Russian disinformation studies report explicit, often aggressive and blatantly 'in your face' tactics by accounts that can rather easily be traced back to a particular Russia-origin network. In our 6 year ongoing study, we observe a subtler, 'smoke and mirrors' tactic by accounts that are (with one specific exception, which is the 'ISIS oil' campaign) several degrees separated from the usual suspect clusters that have been plaguing Western information ecosystems for quite some time. This demonstrates the value of the comparative and longitudinal studies for which we call.

Endnotes

1 Edda Humprecht, Frank Esser, and Peter Van Aelst, "Resilience to Online Disinformation: A Framework for Cross-National Comparative Research," *The International Journal of Press/Politics* 25, no. 3 (July 1, 2020): 493–516, https://doi.org/10.1177/1940161219900126; W Lance Bennett and Steven Livingston, "The Disinformation Order: Disruptive Communication and the Decline of Democratic Institutions," *European Journal of Communication* 33, no. 2 (April 1, 2018): 122–39, https://doi.org/10.1177/0267323118760317.

2 Eleni Kapantai et al., "A Systematic Literature Review on Disinformation: Toward a Unified Taxonomical Framework," *New Media & Society* 23, no. 5 (May 1, 2021): 1301–26, https://doi.org/10.1177/1461444820959296.

3 Barry Buzan and Ole Wæver, "Macrosecuritisation and Security Constellations: Reconsidering Scale in Securitisation Theory," *Review of International Studies* 35, no. 2 (2009): 253–76; Barry Buzan, "Regional Security Complex Theory in the Post-Cold War World," in *Theories of New Regionalism: A Palgrave Reader*, ed. Fredrik Söderbaum and Timothy M. Shaw, International Political Economy Series (London: Palgrave Macmillan UK, 2003), 140–59, https://doi.org/10.1057/9781403938794_8.

4 André Barrinha, "The Ambitious Insulator: Revisiting Turkey's Position in Regional Security Complex Theory," *Mediterranean Politics* 19, no. 2 (May 4, 2014): 165–82, https://doi.org/10.1080/13629395.2013.799353.

5 Ziya Öniş and Şuhnaz Yılmaz, "Turkey and Russia in a Shifting Global Order: Cooperation, Conflict and Asymmetric Interdependence in a Turbulent Region," *Third World Quarterly* 37, no. 1 (January 2, 2016): 71–95, https://doi.org/10.1080/01436597.2015.1086638.

6 Hamid Akin Unver, "The Fog of Leadership: How Turkish and Russian Presidents Manage Information Constraints and Uncertainty in Crisis Decision-Making," *Southeast European and Black Sea Studies* 18, no. 3 (July 3, 2018): 325–44, https://doi.org/10.1080/14683857.2018.1510207.no. 3 (July 3, 2018

7 H. Tarık Oğuzlu, "Turkish Foreign Policy at the Nexus of Changing International and Regional Dynamics," *Turkish Studies* 17, no. 1 (January 2, 2016): 58–67, https://doi.org/10.1080/14683849.2015.1136088.

8 Hamid Akin Unver, "Russian Disinformation Ecosystem in Turkey" (Istanbul: Center for Economic and Foreign Policy Research (EDAM), March 2019).

[9] Keith Johnson Gramer Robbie, "Who Lost Turkey?," *Foreign Policy* (blog), July 19, 2019, https://foreignpolicy.com/2019/07/19/who-lost-turkey-middle-east-s-400-missile-deal-russia-syria-iraq-kurdish-united-states-nato-alliance-partners-allies-adversaries/.

[10] Maria Tsvetkova Kelly Lidia, "Russia Says It Has Proof Turkey Involved in Islamic State Oil Trade," *Reuters*, December 2, 2015, https://www.reuters.com/article/us-mideast-crisis-russia-turkey-idUSKBN0TL19S20151202.

[11] State Department and Pentagon arguments need dissecting in detail. Their argument is that 'Turkey buys oil from ISIS' and 'Turkish oil companies buy oil from ISIS-controlled regions' are two different narratives. Turkish companies predictably buy oil from the same Syrian oilfields they have been trading with since 1970s. When those areas were overtaken by ISIS, oil trade continued, and Turkish tankers continued to carry crude from the same oilfields (since oilfields don't move) they have been operating from for decades. Although this looks like a tiny detail, it is specifically this kind of nuances that feed more sophisticated and successful Russian influence operations. "Pentagon Rejects 'Preposterous' Idea That Turkey Is Aiding ISIS Oil Trade," NBC News, December 2, 2015, https://www.nbcnews.com/video/pentagon-rejects-preposterous-idea-that-turkey-is-aiding-isis-oil-trade-577939523771; Lucas Tomlinson, "State Dept. 'Rejects' Russia's Claims That Turkey Smuggling ISIS Oil," *Fox News*, December 4, 2015, https://www.foxnews.com/politics/state-dept-rejects-russias-claims-that-turkey-smuggling-isis-oil.

[12] Shaun Walker, "Erdoğan and Putin Discuss Closer Ties in First Meeting since Jet Downing," *The Guardian*, August 9, 2016, sec. World news, http://www.theguardian.com/world/2016/aug/09/erdogan-meets-putin-leaders-seek-mend-ties-jet-downing-russia-turkey.

[13] Tim Arango and Ceylan Yeginsu, "Turks Can Agree on One Thing: U.S. Was Behind Failed Coup," *The New York Times*, August 2, 2016, sec. World, https://www.nytimes.com/2016/08/03/world/europe/turkey-coup-erdogan-fethullah-gulen-united-states.html.

[14] Andrew Finkel, "Turkey and Russia Have United over the Karlov Killing. But Deep Tensions Remain | Andrew Finkel," *The Guardian*, December 20, 2016, sec. Opinion, http://www.theguardian.com/commentisfree/2016/dec/20/andrei-karlov-assassination-turkey-russia-putin-erdogan.

[15] Atıf Emre Yüksel, Yaşar Alim Türkmen, Arzucan Özgür, and Berna Altınel. «Turkish tweet classification with transformer encoder.» in *Proceedings of the International Conference on Recent Advances in Natural Language Processing* (RANLP 2019), pp. 1380-1387. 2019.

Middle East Influence Operations:
Observations Across Social Media Takedowns

Renée DiResta, Stanford University; Josh A. Goldstein, Stanford University; and Shelby Grossman, Stanford University

Introduction

The 2009 Green Movement in Iran and 2011 Arab Spring uprisings across the Middle East and North Africa (MENA) region showed governments the power of social media activism and its potential threat to regime stability.[1] Early media coverage and academic research posited that the new platforms would be democratizing.[2] However, in the years that followed, the region's governments transformed from passive targets of social media mobilization to active online agents themselves—shaping and constraining public opinion for their own political ends.

The regimes incorporated social media activities into their own domestic and foreign policy toolkits; social networks became yet another broadcast channel upon which to communicate state messages and transmit propaganda. The affordances of social media enabled a range of novel tactics for covert information operations in particular: profiles absent verification, for example, can be used to create personas to conduct agent-of-influence activities. Because much of the world now has accounts on the largest social media platforms, such as Facebook and Twitter, state actors can target the citizens of other nations directly.[3] For instance, one information operation originating in Iran focused on countries ranging from Bosnia and Bangladesh to Mauritania and Morocco to Senegal and Sudan.[4]

When platforms identify this type of manipulative activity targeting users, they take it down, removing (in the case of Facebook, for example) the Pages as well as the accounts identified as active participants. Assessing these operations is important for a number of reasons: they are happening (or, at least, being detected) with increasing frequency, and they have the potential to destabilize countries or exacerbate geopolitical tensions. The dominant model for thinking about state-sponsored influence operations remains what is sometimes called the "Russian playbook" after the activity attributed to Russia's Internet Research Agency. In their 2014-2017 operation targeting the United States, the Internet Research Agency used fake personas masquerading as citizens of the targeted country, front news media that on the surface appeared to be activist publications, cross-platform deployment of both personas and fronts, and attempts to leverage distinct facets of identity (race, gender, religion etc.) to create tension with those from different demographic or ideological groups.[5] However, studies of information operations initiated by other state actors reveal a breadth of tactics: examination of networks attributed to the Chinese Communist Party, for example, suggests that it does not invest the time or effort required to develop convincing personas, but instead creates new clusters of accounts to address particular topics, often using them as amplifiers rather than as message initiators.[6] Another novel approach appeared in a network attributed to the Pakistani military, which leveraged many real accounts (alongside some fake ones) to create fan Pages for the military. The accounts mass-reported perceived enemies of the government to silence counter-speech.[7] So, the "Russian playbook," if one exists at all,[8] is but one approach.

In this paper we take information operations originating in the MENA region as a class. Since most MENA operations have not targeted Western elections—where much of the research efforts have focused—they remain relatively understudied, as Unver notes in this collection.[9] We ask: what are the trends, tactics, and promoted narratives from the networks disrupted in the MENA takedowns? Our goal is not to conduct new research on specific campaigns, but to contribute to a broader effort to look across existing information operations research for emerging themes and trends.[10] Is there a discernible "playbook" common to individual country or regional political operations?

Data and Methods

To answer these questions, this paper examines a dataset of all known Facebook and Twitter takedowns centered on the MENA region. We built the dataset in three steps. First, we sifted through all takedown announcements by Facebook and Twitter, and identified MENA-centered takedowns. A takedown is considered "MENA-centered" if both the attributed country of origin and at least one target country are from the MENA region.[11] To assess an operation's location of origin, we rely on attributions made by Facebook and Twitter. Assessing target was more difficult; although Facebook and Twitter often include in their takedown announcement the "focus" of an operation, what a network of accounts discusses and who it targets are two distinct questions. Who, for example, is the target of an information operation in Arabic that discusses Libya? It may be Libyans, but it could also be regional governments; information operations frequently try to convince governments to act in a manner favorable to the perpetrator. We estimated the target based on language and the most frequently discussed topics.[12]

Second, we coded key variables for each takedown: actor attribution, target audiences, platforms used, size of the network, number of followers amassed, narratives promoted, and tactics used to advance those narratives. To assess these variables, we relied on platform and researcher takedown reports for individual operations. Since August 2018, Facebook[13] and Twitter[14] have announced dozens of takedowns with brief summaries of the influence operation activity and reason for removal. The platforms also partner with third-party cybersecurity research firms and academic institutions, which publish deeper independent analysis of the networks.[15] Coding some attributes—like the number of accounts removed in the takedown—was straightforward, while coding other variables—like tactics—was more challenging. If a social media platform or a researcher report referenced that a tactic was used, we concluded that an operation included the aforementioned tactic. However, the lack of mention by a social media platform or researcher report does not necessarily imply that a tactic was not used; our coding on tactics should thus be considered a lower-bound of total uses.

Third, in cases where researcher reports were not available, we looked at the hashed Twitter takedowns directly.[16] This approach was not possible for Facebook takedowns without researcher reports, as Facebook does not provide a public archive of information removed in takedowns.

The result is a dataset of 46 information operations, originating from ten MENA countries, whose removals were announced by Facebook or Twitter between August 2018 and March 2021.

While comprehensive in terms of MENA takedowns announced by Facebook and Twitter, our data set is not necessarily representative of all information operations in the MENA region, as shown by Jones in his contribution to this collection. Some operations may have evaded detection; others may not have been disclosed to the public.[17]

Iran was the most frequent country of origin; 20 of the 46 takedowns in the dataset originated from Iran. Egypt was second (10 takedowns), the UAE third (6), and Saudi Arabia fourth (5).[18] The coding of country of origin does not mean that the operation was directed by the government in that particular country, but simply that Facebook or Twitter reported that the accounts originated from the country in question.

Attribution

We examined attributions made by Facebook and Twitter for each Middle East takedown, shown in Table 1.[19] Among the 46 takedowns, 24% were linked to a government and 26% were attributed to a marketing, PR, or IT firm. The use of marketing firms is not unique to the Middle East. Governments increasingly outsource influence operations to digital mercenaries because of access to external expertise and plausible deniability.[20] Around half of the takedowns attributed to marketing firms involve the UAE or Egypt.

Table 1: Summary of Middle East Takedowns by Facebook and Twitter, Coded by Attribution

Type of Entity Involved	Number of Takedowns	Location of Origin	Attribution from Platform (T=Twitter, F=Facebook; Data of Public Disclosure)
Unspecified Individuals	18	Iran	• Iran: "linked to the network we removed in October 2020" (F: April 6, 2021) • "individuals in Iran with academic backgrounds" (F: March 3, 2021) • "individuals in Tehran" (F: March 3, 2021) • "links to individuals in Iran" (F: January 12, 2021) • "originated in Iran" (F: October 21, 2019) • "originated in Iran" (F: May 28, 2019) • "our review linked these accounts to Iran" (F: March 26, 2019) • "tied to Iran" (F: January 31, 2019) • Iran, not specified (F: August 21, 2018 Iran Network 3) • "originating in Iran" (T: August 21, 2018)21 • "may have origins in Iran" (T: January 31, 2019) • "operating from Iran" (T: October 8, 2020)22
		Morocco	• "originated primarily in Morocco" (F: March 3, 2021)
		Palestine, UAE	• "individuals in Palestine and UAE"23 (F: February 9, 2021)
		Yemen	• "originating in Yemen" (F: August 6, 2020)
		Iraq	• "in Iraq" (F: September 16, 2019)
		UAE	• "operating uniquely from the UAE" (T: September 20, 2019)24
		Saudi Arabia	• "associated with Saudi Arabia" (T: April 2, 2020)25
Marketing, PR, or IT Firm	12	Egypt	• "Bee interactive, a marketing firm in Egypt" (F: April 6, 2021) • "Maat, a marketing firm in Egypt" (F: April 2, 2020) • "two marketing firms in Egypt, New Waves and Flexell" (F: March 2, 2020)
		Palestine, UAE, Belgium	• "a recently created marketing firm called Orientation Media in Belgium"26 (F: February 9, 2021)
		Morocco	• "Qualitia Systems, a marketing firm in Morocco, also known as Marketing Digital Maroc" (F: January 12, 2021)
		Iran	• "linked to individuals associated with EITRC, a Tehran-based IT company" (F: November 5, 2020)
		Israel	• "Israeli commercial entity, Archimedes Group" (F: May 16, 2019)
		Tunisia	• "a Tunisia-based PR firm Ureputation" (F: June 5, 2020)
		UAE, Egypt, Nigeria	• "Charles Communications in UAE, MintReach in Nigeria and Flexell in Egypt" (F: October 3, 2019)
		UAE, Egypt	• "New Waves in Egypt, and Newave in the UAE" (F: August 1, 2019) • "created and managed by DotDev, a private technology company operating in the UAE and Egypt" (T: September 20, 2019)27
		Saudi Arabia	• "Smaat, a social media marketing and management company based in Saudi Arabia" (T: December 20, 2019)28

Type of Entity Involved	Number of Takedowns	Location of Origin	Attribution from Platform (T=Twitter, F=Facebook; Data of Public Disclosure)
Government-Linked	11	Iran	• "individuals associated with the Iranian government" (F: October 27, 2020) • "Islamic Republic of Iran Broadcasting Corporation." (F: May 5, 2020) • "Iranian state media" (F: August 21, 2018 Iran Network 1) • "all are associated with — or directly backed by — the Iranian government" (X3) (T: August 21, 2018)29
		Kurdistan	• "Zanyari Agency, part of the intelligence services of the Kurdistan Regional Government in Iraqi Kurdistan" (F: June 5, 2020)
		Saudi Arabia	• "individuals associated with the government of Saudi Arabia" (F: August 1, 2019) • "linked to Saudi Arabia's state-run media apparatus" (T: September 20, 2019)30 • "with ties to the Saudi government" (T: October 8, 2020)31
		Egypt	• "El Fagr network...Information we gained externally indicates it was taking direction from the Egyptian government." (T: August 2, 2020)32
Third Party News	3	Israel	• "ElBaladd, a news website in Israel" (F: April 6, 2021)
		Egypt	• "an Egyptian newspaper El Fagr" (F: October 3, 2019)
		Iran	• "Liberty Front Press" (F: August 21, 2018 Iran Network 2)
Political Group	2	Albania	• "MEK, an exiled militant opposition group from Iran now based in Albania" (F: April 6, 2021)
		Egypt, Turkey, Morocco	• "individuals in Egypt, Turkey and Morocco associated with the Muslim Brotherhood" (F: November 5, 2020)
TOTAL	**4733**		

Goals and Narratives

After assessing the actors to which the operations were attributed, we examined the content shared by these networks to understand the potential goals of the operation. There were four primary objectives spanning all operations:

- Attempts to cast one's own government, culture, or policies in a positive light
- Advocacy for or against specific policies
- Attempts to make allies look good and rivals look bad to third-party countries
- Attempts to destabilize foreign relations or domestic affairs in rival countries

In this section, we describe these objectives, and offer examples of narratives leveraged in the effort to achieve them.

Promoting (and Protecting) One's Image

The networks engaged in promoting a positive image of their country of origin amplified narratives that cast their leadership and policies as beneficial to both their own citizens, and often the broader region as well. Data sets attributed to Iran-linked actors contained content

that positioned Iran as the champion of the oppressed and the leader of the Muslim world. Iran-based networks often championed Palestinian rights and denounced US, Israel, and Saudi Arabia's regional interventions and collusion. Iran was portrayed as a bulwark against neocolonialism and the "West," and a stabilizing force in the region. Showcasing Iran's capability to stand up for the oppressed and confront the "West," posts from other Iran-linked networks also boasted of the Iranian military's threat to Israel and the United States.[34] Furthermore, at least two takedowns had assets promoting the Iranian Supreme Leader's religious teachings, potentially to increase his appeal among Muslims outside Iran and further Iran's cultural diplomacy.[35] Accounts that tweeted in many languages reveal Iran's attempt to promote its appeal among a global audience. Such messages are consistent with transparently-attributable, "white propaganda" Iranian state media narratives.[36] However, the covert influence campaigns additionally allowed accounts with seemingly no visible ties to Iran to launder Iranian propaganda to unsuspecting users.[37] Examining a takedown linked to the Islamic Republic of Iran Broadcasting Corporation, Graphika researchers noted, "Many of its assets conducted what could have been considered classic public diplomacy, if it had been done overtly: promoting Iran's successes and spiritual authority to Arabic- and English-speaking audiences."[38]

Saudi Arabia largely promoted its achievements for its domestic population, contrasting its successes with Iran's domestic failures. It also sought to present an attractive image of Saudi Arabia to the Western world. For instance, an August 2019 takedown attributed to individuals associated with the Saudi government promoted Saudi Arabia's military and social achievements.[39] The network promoted the successes of the Saudi Armed Forces, and the Crown Prince Mohammad bin Salman's economic and social reform plan, "Vision 2030." One Twitter takedown portrayed the Saudi Crown Prince as personable and relatable, showing him trying VR games and leading a traditional dance.[40] Some of the network's posts promoted the country's progress in women's rights, featuring Saudi women who were pushing traditional boundaries as horseback racers, top chefs, and more. These successes were implicitly accredited to the government. This Saudi network also featured other feel-good posts highlighting national points of pride.

The Saudi networks also engaged in reputational damage control during controversies that garnered significant media coverage, such as when Jeff Bezos' phone was hacked or after the journalist Jamal Khashoggi was murdered in the Saudi consulate in Istanbul in 2018.[41] The takedown data set revealed that the networks tried different strategies for deflecting blame of the Khashoggi murder from Saudi Arabia, offering a range of overlapping (and at times conflicting) narratives that ranged from denying the murder, to claiming it occurred elsewhere, to attacking Khashoggi's character.[42]

As with the Saudi campaigns, the campaigns that originated in the UAE showcased the country's social and economic achievements. Campaigns originating in the UAE attempted to create the perception of broad, widespread global praise for the country. A September 2019 Twitter takedown of 4,248 accounts operating uniquely from the UAE exemplifies this.[43] A set of accounts claimed to be of diverse nationalities and posted praise of the UAE in languages including Arabic, Chinese, English, French, German, Hebrew, Italian, Japanese, Korean, Persian, Polish, Portuguese, Russian, Spanish, and Turkish. Many of the accounts posted about a visit that Pope Francis made to the UAE, and emphasized that the UAE is a tolerant country. Others promoted the UAE as an attractive tourist destination (e.g. "Summer goals #UAE #SaturdayMorning" from a Twitter account that purported to belong to an Australian activist). They touted the UAE's international events, such as the World Government Summit, its celebration of Chinese New Year, and the Abu Dhabi International Triathlon.

Manufacturing Consensus For or Against Specific Policies

Astroturfing[44] accounts were used to create an impression of domestic grassroots support or opposition not only to certain governments, but also to particular government policies.

For example, we saw astroturfing related to the Iran-Russia Defense Agreement: In an April 2020 Twitter takedown, a network of accounts associated with Saudi Arabia created the impression of local Iranian opposition to a potential joint defense agreement between Russia and Iran.[45] Amidst rising U.S.-Iran tensions in the summer of 2019, reports surfaced that Iran was pursuing a joint defense agreement with Russia. Posing as Iranians, the accounts used the English hashtag #GetLostFromIranRussia. They portrayed the agreement as Russian colonial intervention infringing on Iran's sovereignty.

We also observed astroturfing against the Grand Ethiopian Renaissance Dam Project. A March 2021 Facebook takedown attributed to Bee Interactive, a marketing firm in Egypt promoted domestic resistance to the project, which threatened Egypt's fresh water supplies. Five Pages posed as independent news outlets and criticized the dam.[46]

Denigrating Regional Rivals

Regional rivalries, often between a Saudi Arabia/UAE/Egypt axis on one hand, and an Iran/Turkey/Qatar axis on the other, feature prominently in MENA takedowns.[47]

Networks attributed to the former set commonly portrayed the latter set as having a destabilizing presence in the region. Among the 15 takedowns that originated in Egypt, Saudi Arabia, and/or the UAE, at least 10 portrayed Qatar, Turkey, and/or Iran as sponsors of terrorism. In addition to underscoring the rivals' destabilizing regional activities, the campaigns also criticized their domestic performance, potentially to reduce the rivals' regional appeal or domestic stability. They showcased moral corruption in Iran, economic collapse in Turkey, and human rights violations in Qatar, frequently leveraging accounts posing as locals in the target country to push those narratives.

Isolating Qatar was by far the most salient example of MENA takedown networks trying to denigrate rivals. In June 2017, Saudi Arabia and other Arab countries cut diplomatic ties and launched a land, air, and sea blockade against Qatar. The intra-Gulf crisis reflected a geostrategic and ideological gulf between Qatar and the blockading countries. The latter issued a list of demands that Qatar had to meet for the blockade to end.[48] These included demands that Qatar curb its relations with Iran and Turkey, sever ties to Islamist and terrorist groups, and shut down its state-funded broadcaster, Al Jazeera, and affiliate stations. The blockade lasted three and a half years until a January 2021 GCC summit aimed at reconciliation. During the blockade, narratives attempting to isolate Qatar from any potential allies and create rifts featured in the takedown data sets:

- **Qatar-United States Rift:** Some posts sought to create a wedge between Qatar and the United States. Accounts relentlessly portrayed Qatar as a sponsor of terrorism. One Page, from an October 2019 takedown attributed to three commercial firms, promoted the narrative that Qatar indirectly supported the 9/11 attacks against the United States.[49]

- **Qatar-Turkey Rift**: Assets in the October 2020 Twitter takedown linked to the Saudi government posted or retweeted posts that Turkey killed and insulted several members of Qatar's royal family, that Turkey was occupying Qatar, and that "Erdogan is used to exploiting the young #Qatari Emir."[50]

- **Qatar-Iran Rift**: For example, a February 2020 Facebook takedown attributed to two marketing firms in Egypt promoted allegations that Qatar played a role in the assassination of the nationally popular Iranian general, Qassim Soleimani.[51]

The rift-creating narratives are in line with the blockading countries' geopolitical interests. Qatar's ties with Iran and Turkey were not only stimuli for the blockade. They also undermined the blockade's aim of severely pressuring Qatar, as Iran and Turkey provided critical exports to Qatar during the blockade. Iran and Turkey, too, appeared in the narratives claiming that their leaders support terror groups, including ISIS and the Muslim Brotherhood.

The networks also pushed claims of Qatar, Turkey, and Iran's interference in particular countries. An August 2019 Facebook takedown attributed to two marketing firms –

New Waves in Egypt, and Newave in the UAE – amplified claims that Qatar and Turkey support terrorist groups in Africa and the Middle East.[52] The takedown particularly amplified narratives that Qatar was involved in a terror attack in Somalia.[53] It used fake news outlets to amplify such reports. One Page, posing as the social media arm of the website Somalianow, also promoted its articles criticizing Qatar's investments in Africa.

Iranian operations frequently criticized regional rivals. Of the 20 takedowns originating in Iran, at least 15 criticized Israel, Saudi Arabia, and/or the UAE. The takedowns commonly portrayed these countries as corrupt and complicit in Western crimes, or as un-Islamic. Their operations also painted a picture of Western neocolonialism in the region, abetted by rival regional governments like Saudi Arabia and the UAE. An April 2020 Facebook takedown attributed to the Islamic Republic of Iran Broadcasting Corporation (IRIB) likewise promoted narratives of the Saudi royal family's corruption. For instance, one Arabic-language Page, largely active between 2014 and 2015, was named "Saudi opposition and free speech page." Its 'About' section described itself as a "Private page of Saudi revolutionaries that transmits the truth to the outside and to anyone who is looking for the truth, in order to free the country [Saudi Arabia] from the [House of] Saud, may they be cursed by God." Its memes and texts portrayed the Saudi ruling family as a puppet regime serving the United States and Israel.

Destabilizing Rival Governments

Denigrating regional rivals extended beyond policy criticism and into outright attempts to undermine leaders and destabilize regimes.

At least two takedowns promoted independence for Somaliland. The August 2019 Facebook takedown attributed to New Waves in Egypt and Newave in the UAE frequently posted on this topic.[54] Five accounts from the April 2020 Twitter takedown associated with Saudi Arabia claimed to be based in Somaliland. Their tweets extolled the wildlife, nature, and physical beauty of Somaliland. Some mentioned the "rebirth of Somaliland," and claimed they were "proud to be #Somalilanders." Accounts also tweeted the English hashtag, #Somaliland_not_somalia.[55]

The campaigns promoting Somaliland's independence are an extension of Saudi Arabia and the UAE's geopolitical interests. During the intra-GCC crisis, Saudi Arabia and the UAE reportedly pressured Somalia's newly elected president, Mohamed Abdullahi Mohamed (also referred to as Farmajo), to sever ties with Qatar. Farmajo insisted on remaining neutral. However, reports that the president received funds from Qatar ahead of his election, and his appointment of officials close to Doha, raised Abu Dhabi's doubts of his neutrality.

This delegitimization strategy was deployed against other governments as well:

- **Sudan:** The April 2020 El Fagr takedown promoted narratives undermining the Sudanese government. Amidst protests in Sudan in June 2019, many fake accounts supported the protestors, saying the protesters were rejecting the Muslim Brotherhood.[56]

- **Syria:** The April 2020 Twitter takedown associated with Saudi Arabia advanced narratives of the Syrian regime's domestic unpopularity. Thirty-six accounts had Syria-related usernames or references to Syria in their profile. Their tweets criticized Syria's President Bashar al-Assad.[57]

- **Libya:** Networks often attacked Fayez al-Sarraj, the former Prime Minister of Libya's Government of National Accord (GNA). A Saudi-linked Twitter takedown amplified the hashtag "Sarraj the traitor of Libya" (translated from Arabic). Feeding off this inauthentically widespread use of the hashtag, a pro-Haftar YouTube channel[58] and several articles[59] published about the "trending" hashtag; they asserted that many Libyans commonly viewed Sarraj's agreement with Erdogan as a betrayal of their country.[60]

Iran: In at least one takedown, accounts boldly criticized Iran's Supreme Leader Ayatollah Ali Khamenei, and the late IRGC Quds Force commander, Qassem Soleimani. In

addition to employing hashtags like #IranRegimeChange (translated into English), the suspended accounts promoted candidates for the incumbent regime's replacement. They used the pro-Iranian monarchy hashtag #G20RecognizePahlavi and promoted the Mojahedin-e Khalq (MEK).[61]

Tactics

In our coding process we made note of tactics, techniques, and procedures that appeared within research reports describing each operation. Around three quarters of the takedowns included assets that claimed to be news outlets. Influence operations involving accounts masquerading as media outlets is an extremely common, recurring approach. However, there is some nuance involved in how manipulators execute on this approach. In the Middle East data set, the "news outlet" accounts took on a range of forms:

- **Leveraging quasi-real slanted news outlets**. For example, two takedowns were attributed to the El Fagr newspaper in Egypt. While El Fagr claims to be an "independent weekly newspaper,"[62] Twitter noted that information "gained externally indicates it was taking direction from the Egyptian government."[63]

- **Creating front media**. Most news outlets linked to suspended social media operations in the dataset fell into this category. They often reposted content from other news sites, selectively publishing news stories that advanced the network's objectives. The accounts' occasional original content was often poorly written, replete with grammar and spelling mistakes. The fake news outlets used standard newspaper naming conventions, like "Sudan Today" and "Afghan Mirror."

- **False franchising.** Some Facebook Pages covered in takedown reports were designed to look like regional branches of authentic large news outlets. For example, one asset posed as the regional Page for the Huffington Post, calling itself the Huffpost Taounatepress.[64]

- **Impersonating real news outlets.** Some takedowns impersonated real news organizations. For instance, in a May 2019 Iran takedown, one Page, @AlArabyi, impersonated the Saudi-funded news channel @AlArabiya. The Arabic-language mimicked the name, logo, and visual branding of AlArabiya.net.[65] Occasionally these outlets used "typo-squatting," employing Facebook URLs that mimic the URLs for authentic media outlets, with minor typographical changes to trick users into mistaking them for the authentic domains. A November 2020 takedown attributed to individuals in Iran and Afghanistan engaged in typo-squatting of a popular Facebook Page; the Page facebook.com/aff.varzeshi spoofed the genuine facebook.com/aff.varezshi by reversing the "ez."[66]

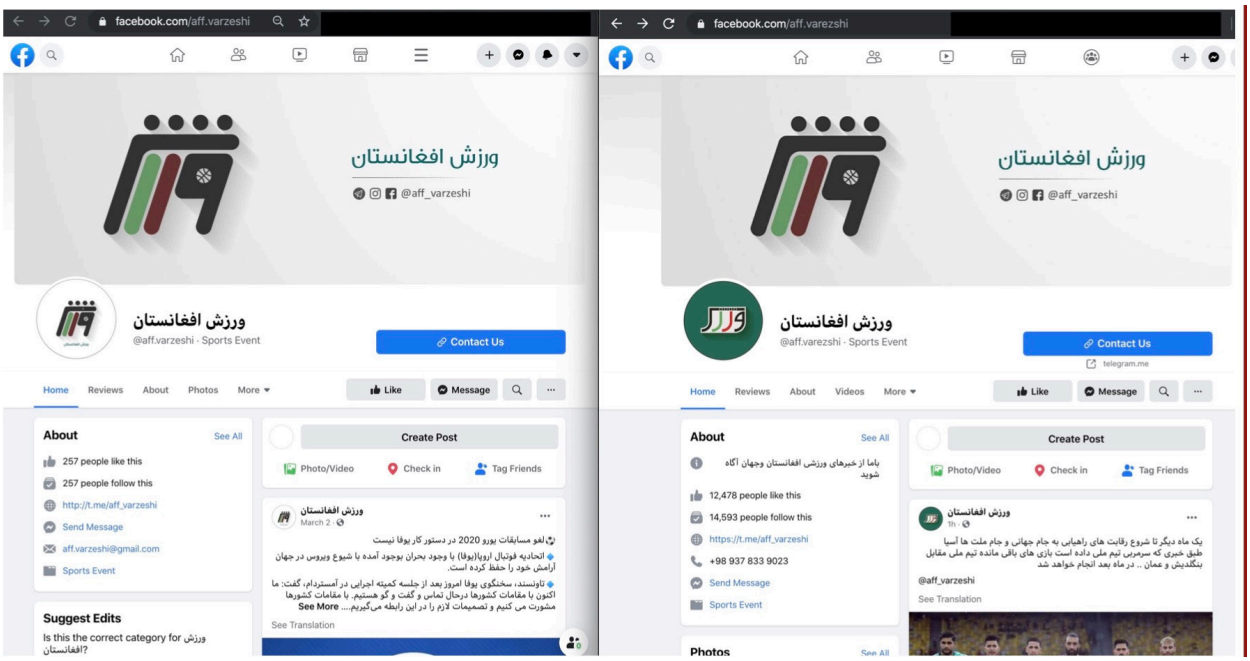
We observed a number of tactics to build audiences and increase account legitimacy. These included:

- **Non-political, humorous or fashion content**. Gathering a relatively large and broad following with light-hearted content, the assets would then occasionally redirect their viewers to the more narrowly politicized assets in the network.[67] Sometimes these two roles – posting engaging content and advancing political narratives – were conducted by the same asset. They interspersed strategic political posts amidst filler content.

- **Handle-switching**, where an account grows its following, perhaps with spammy follow-back behavior, then deletes its old tweets and changes its handle. We saw this tactic used by an account that, once it established a following, pretended to be an interim Qatari government.[68]

- **Early account creation dates**, potentially from hacking or purchasing real accounts.

- **Bolstering accounts on one platform with "off-platform" presence** elsewhere online, for example by having a fictitious persona publish op-eds.[69]



The authentic facebook.com/aff.varezshi Page, right, and the mimicking facebook.com/aff.varzeshi Page, left.[70]

- **Astroturfing**. Creating accounts that posed as locals in target countries.[71] The tactic of using 'fake local' accounts may be manifold. It may be to distort the picture for journalists and analysts who rely on social media to gauge public sentiment in the subject country. It may also be to influence real locals in the target country to overestimate the predominance of a particular public sentiment, which might in turn quiet the voices of those harboring opposing viewpoints, or intensify the attitudes of those sharing those views.[72] Networks advocating protests and questioning a government's legitimacy may seek to foment domestic overthrow of unfriendly governments; the 2011 Arab Spring made clear the power of social media as a tool to spur popular uprising. However, the networks' ability to achieve those potential objectives is debatable; many of these accounts received little engagement and researchers have long struggled to measure the impact of influence operations.[73]

Marc Owen Jones' article in this collection discusses an important recent tactical innovation, particularly on Twitter: the use of chopped hashtags. With chopped hashtags, "sock puppet accounts dilute and pollute critical hashtags using abbreviated versions of the real hashtags." While we have not seen this tactic in the takedowns analyzed in this paper, we expect to see this tactic in future takedowns.

Conclusion

Assessing attributed influence operation narratives and methods offers visibility into what topics state and non-state actors have chosen to prioritize, and potentially can elide a set of recurring tactics, techniques, and procedures to help attribute subsequent operations. However, in the culmination of our cross-dataset analysis of MENA region takedowns, we observed a very wide breadth of tactics and narratives even within operations attributed to one single state actor. Perhaps the best way to describe the "Middle East playbook" is that many information operations attributed to countries in the Middle East use a "kitchen sink" approach. Across multiple individual takedowns we saw a wide array of tactics and narratives employed in service to multiple geopolitical objectives – far more than, for example, activity observed in Chinese operations in which clusters of accounts are largely

dedicated to producing or amplifying content about a single self-promotional narrative, with a recurring focus on bolstering China's own image. The huge breadth of issues we see covered in these MENA region operations suggests, perhaps, that covert social media influence operations are becoming a normal part of propaganda and influence efforts by these regimes rather than something reserved for events or topics of extremely high impact or significance. The narratives the networks promoted often systematically aligned with the geopolitical interests of their country of origin. However, the social media campaigns' efficacy is debatable. Most of the Middle East takedowns achieved low engagement. Moreover, the causal mechanisms between exposure and changes in attitudes and behavior have yet to be strongly established. Nonetheless, investigating these information operations helps to better understand regional and domestic politics in the Middle East and North Africa.

Endnotes

[1] Philip N. Howard and Muzammil M. Hussain. 2011. "The Upheavals in Egypt and Tunisia: The Role of Digital Media." *Journal of Democracy* 22(3): 35-48.

[2] See for example: McGarty, Craig, Emma F. Thomas, Girish Lala, Laura GE Smith, and Ana- Maria Bliuc. 2014. "New Technologies, New Identities, and the Growth of Mass Opposition in the Arab Spring." *Political Psychology* 35(6): 725-740.

[3] For example, across five Middle Eastern and North African countries surveyed in a 2018 Pew study, 68% of respondents said that they use social networking sites. Poushter, Jacob, Caldwell Bishop, and Hanyu Chwe. 2018. "Social Media Use Continues to Rise in Developing Countries but Plateaus Across Developed Ones." *Pew Research Center.*

[4] https://about.fb.com/wp-content/uploads/2020/05/April-2020-CIB-Report.pdf

[5] Renée, DiResta, Kris Shaffer, Becky Ruppel, David Sullivan, Robert Matney, Ryan Fox, Jonathan Albright, and Ben Johnson. 2019. "The Tactics & Tropes of the Internet Research Agency." *New Knowledge Report.*

[6] Renée DiResta, Carly Miller, Vanessa Molter, John Pomfret, and Glenn Tiffert. 2020. "Telling China's Story: The Chinese Communist Party's Campaign to Shape Global Narratives." *Hoover Institution/Stanford Internet Observatory Report.* Accessed: https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/sio-china_story_white_paper-final.pdf.

[7] https://cyber.fsi.stanford.edu/io/news/reporting-duty

[8] For a thoughtful critique on the notion that a single "Russian Playbook" exists, see: François, Camille. "Moving Beyond Fears of the 'Russian Playbook.'" *Lawfare.* September 15, 2020. Accessed: https://www.lawfareblog.com/moving-beyond-fears-russian-playbook

[9] For one data point, the Partnership for Countering Influence Operations at the Carnegie Endowment for International Peace created a catalogue of 460 counter-influence operations initiatives. The project found that 44% of initiatives studying influence operations are located in the North America and 37% in Europe, perhaps partially explaining the over-indexing on elections in those regions. See: Smith, Victoria. "Mapping Worldwide Initiatives to Counter Influence Operations." *Carnegie Endowment for International Peace*. December 14, 2020. Accessed: https://carnegieendowment.org/2020/12/14/mapping-worldwide-initiatives-to-counter-influence-operations-pub-83435

[10] For other recent comparative analyses, see: Bradshaw, Samantha, Hannah Bailey, and Philip N. Howard. 2021. "Industrialized Disinformation: 2020 Global Inventory of Organized Social Media Manipulation." *Computational Propaganda Research Project*. Accessed: https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/127/2021/01/CyberTroop-Report-2020-v.2.pdf; Goldstein, Josh A. and Shelby Grossman. "How disinformation evolved in 2020." *Lawfare.* January 4, 2021. Accessed: https://www.brookings.edu/techstream/how-disinformation-evolved-in-2020/; Martin, Diego A., Jacob N. Shapiro, and Julia Ilhardt. 2020. "Trends in Online Influence Efforts." *Working Paper.*

[11] In this paper, we consider MENA countries to consist of Algeria, Bahrain, Egypt, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Qatar, Saudi Arabia, Syria, Tunisia, United Arab Emirates, and Yemen.

[12] After we completed our research, Facebook unrelatedly published a dataset of all takedowns from its platform. We therefore checked our coding against their attribution and target coding. Twitter, however, has not yet published a similar spreadsheet. See: Gleicher, Nathaniel, Margarita Franklin, David Agranovich, Ben Nimmo, Olga Belogolova, and Mike Torrey. May 2021. "Threat Report: The State of Influence Operations 2017-2020." *Facebook*. Accessed: https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf

[13] https://about.fb.com/news/tag/coordinated-inauthentic-behavior/

[14] https://transparency.twitter.com/en/reports/information-operations.html

[15] https://about.fb.com/news/2018/11/investigating-threats/#working-with-partners

[16] https://transparency.twitter.com/en/reports/information-operations.html

[17] As Facebook noted in one of its takedown reports, "We routinely take down less sophisticated, high-volume inauthentic behaviors like spam and we do not announce these enforcement actions when we take them" https://about.fb.com/wp-content/uploads/2020/05/April-2020-CIB-Report.pdf

[18] Some takedowns originated in multiple countries. For instance, one Twitter takedown announced in April 2020 and associated with Saudi Arabia operated out of three countries: Saudi Arabia, Egypt, and the UAE. https://twitter.com/TwitterSafety/status/1245682443241259010?s=20

[19] For Twitter operations, we footnote to Twitter's announcement of the operation that includes its attribution language. We do not include Facebook citations since Facebook itself now provides a single spreadsheet with links to each takedown announcement, which can be found here: https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf p. 43.

[20] Shelby Grossman and Khadeja Ramali. "Outsourcing Disinformation." *Lawfare.* December 13, 2020.
[21] https://twitter.com/TwitterSafety/status/1032055161978585088?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1032055161978585088&ref_url=https%3A%2F%2Ftechcrunch.com%2F2018%2F08%2F21%2Ffacebook-and-twitter-remove-hundreds-of-accounts-linked-to-iranian-and-russian-political-meddling%2F%3B%20https%3A%2F%2Fblog.google%2Ftechnology%2Fsafety-security%2Fupdate-state-sponsored-activity%2F
[22] https://blog.twitter.com/en_us/topics/company/2020/disclosing-removed-networks-to-our-archive-of-state-linked-information
[23] We double count this operation as both an example of an attribution to unspecified individuals and a marketing, PR, or IT firm. For more information, see footnote 34.
[24] https://blog.twitter.com/en_us/topics/company/2019/info-ops-disclosure-data-september-2019
[25] https://twitter.com/TwitterSafety/status/1245682443241259010?s=20
[26] We double count this operation as both an example of an attribution to unspecified individuals and a marketing, PR, or IT firm. For more information, see footnote 34.
[27] https://blog.twitter.com/en_us/topics/company/2019/info-ops-disclosure-data-september-2019
[28] https://blog.twitter.com/en_us/topics/company/2019/info-ops-disclosure-data-september-2019
[29] https://blog.twitter.com/en_us/topics/company/2019/information-ops-on-twitter
[30] https://blog.twitter.com/en_us/topics/company/2019/info-ops-disclosure-data-september-2019
[31] https://blog.twitter.com/en_us/topics/company/2020/disclosing-removed-networks-to-our-archive-of-state-linked-information
[32] https://twitter.com/TwitterSafety/status/1245682431975460864?s=20
[33] We double counted one takedown due to mixed attribution categories, so the total number of takedowns in the dataset is 46. The operation we double counted was attributed by Facebook as follows: "Our investigation found links to individuals in Palestine and UAE, in addition to links between a small portion of this network and individuals associated with a recently created marketing firm called Orientation Media in Belgium." We thus code Orientation Media in Belgium under the 'Marketing, PR, or IT Firm' category, and individuals in Palestine and UAE under the 'Unspecified Individuals' category. For more information, see: https://about.fb.com/news/2021/02/january-2021-coordinated-inauthentic-behavior-report/.
[34] https://about.fb.com/news/2019/03/cib-iran-russia-macedonia-kosovo/
[35] https://public-assets.graphika.com/reports/graphika_report_irib_takedown.pdf; https://medium.com/dfrlab/facebook-removes-iran-based-assets-again-f17358ef21f.
[36] See also: Brooking, Emerson T. and Suzanne Kianpour. 2020. "Iranian Digital Influence Efforts: Guerrilla Broadcasting for the Twenty-First Century." *Atlantic Council Report*.
[37] https://public-assets.graphika.com/reports/graphika_report_irib_takedown.pdf
[38] https://public-assets.graphika.com/reports/graphika_report_irib_takedown.pdf
[39] https://about.fb.com/news/2019/08/cib-uae-egypt-saudi-arabia/; https://medium.com/dfrlab/royally-removed-facebook-takes-down-pages-promoting-saudi-interests-edc0ce8b972a
[40] https://blog.twitter.com/en_us/topics/company/2019/info-ops-disclosure-data-september-2019
[41] https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/20200402_blame_it_on_iran_qatar_and_turkey_v2_0.pdf
[42] https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/20200402_blame_it_on_iran_qatar_and_turkey_v2_0.pdf
[43] https://blog.twitter.com/en_us/topics/company/2019/info-ops-disclosure-data-september-2019
[44] The term "astroturf" refers to activities that appear to be grassroots activism from ordinary people, but are in fact paid for or executed by government, institutional, or corporate entities.
[45] https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/20200402_blame_it_on_iran_qatar_and_turkey_v2_0.pdf p. 29.
[46] https://public-assets.graphika.com/reports/graphika_report_inauthentic_beehavior.pdf; https://about.fb.com/wp-content/uploads/2021/04/March-2021-CIB-Report.pdf, p17.
[47] We note that these axis groupings are broad generalizations. The countries' alignments and relations pivot with respect to different issues. Nonetheless, these groupings helpfully depict the broad trends observed across the takedowns.
[48] https://www.theatlantic.com/news/archive/2017/06/saudi-led-blocs-list-of-demands-to-end-qatar-crisis/531408/
[49] https://medium.com/dfrlab/uae-facebook-pages-targeted-qatar-and-muslim-brotherhood-8aec916fa1f7
[50] https://github.com/stanfordio/publications/blob/main/twitter-SA-202009.pdf, p19.
[51] https://about.fb.com/wp-content/uploads/2020/03/February-2020-CIB-Report.pdf
[52] https://medium.com/dfrlab/facebook-disabled-assets-linked-to-egypt-and-uae-based-firms-a232d9effc32
[53] https://www.nytimes.com/2019/07/22/world/africa/somalia-qatar-uae.html
[54] https://about.fb.com/news/2019/08/cib-uae-egypt-saudi-arabia/
[55] https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/20200402_blame_it_on_iran_qatar_and_turkey_v2_0.pdf
[56] https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/egypt_whitepaper.pdf
[57] https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/20200402_blame_it_on_iran_qatar_and_turkey_v2_0.pdf
[58] https://www.youtube.com/watch?v=c_4iA8XMUeg
[59] E.g. https://www.mobtada.com/details/892394
[60] https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/20200402_blame_it_on_iran_qatar_and_turkey_v2_0.pdf
[61] https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/20200402_blame_it_on_iran_qatar_and_turkey_v2_0.pdf
[62] https://www.linkedin.com/company/el-fagr-newspaper/
[63] https://twitter.com/TwitterSafety/status/1245682431975460864
[64] https://about.fb.com/news/2021/01/december-2020-coordinated-inauthentic-behavior-report/
[65] https://medium.com/dfrlab/facebook-removes-iran-based-assets-again-f17358ef21f
[66] https://github.com/stanfordio/publications/blob/main/20201105%20iran%20afghanistan%20report%20v2.pdf
[67] Scholars have historically called this pre-propaganda: propaganda that is not directly related to the political message of the propagandist. See: Golovchenko, Yevgeniy, Cody Buntain, Gregory Eady, Megan A. Brown, and Joshua A. Tucker. 2020. "Cross-Platform State Propaganda: Russian Trolls on Twitter and YouTube during the 2016 U.S. Presidential Election." *The International Journal of Press/Politics* 25(3): 357-389.

[68] https://github.com/stanfordio/publications/blob/main/twitter-SA-202009.pdf

[69] https://about.fb.com/news/2021/03/february-2021-coordinated-inauthentic-behavior-report/

[70] https://github.com/stanfordio/publications/blob/main/20201105%20iran%20afghanistan%20report%20v2.pdf

[71] https://about.fb.com/news/2021/03/february-2021-coordinated-inauthentic-behavior-report/; For a discussion of astroturfing and its potential effects, see: Keller, Franziska B., David Schoch, Sebastian Stier, and JungHwan Yang. 2020. "Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign." Political Communication 37(2): 256-280; Zerback, Thomas, Florian Töpfl, and Maria Knöpfle. 2021. "The disconcerting potential of online disinformation: Persuasive effects of astroturfing comments and three strategies for inoculation against them." *new media & society* 23(5): 1080-1098.

[72] Research on the 'spiral of silence' suggests that people are less likely to share their views when they believe that their views are not widespread. See, for example: Hampton, Keith, Lee Rainie, Weixu Lu, Maria Dwyer, Inyoung Shin, and Kristen Purcell. "Social Media and the 'Spiral of Silence.'" *Pew Research Center*. August 26, 2014. Accessed: https://www.pewresearch.org/internet/2014/08/26/social-media-and-the-spiral-of-silence/; Nielle-Neumann, Elisabeth. "The Spiral of Silence: A Theory of Public Opinion." *Journal of Communication* 24(2): 43-51.

[73] Ben Nimmo. 2020. "The Breakout Scale: Measuring the Impact of Influence Operations." *Brookings Institution.* Accessed: https://www.brookings.edu/wp-content/uploads/2020/09/Nimmo_influence_operations_PDF.pdf

Changing Sources: Social Media Activity During Civil War

Anita Gohdes, Hertie School, Berlin and Zachary C. Steinert Threlkeld, University of California, Los Angeles[1]

Introduction

From isolated protests to country-wide uprisings or organized armed conflict, non-state and government actors have learned that their actions are likely to be caught on cellphone camera, and that controlling the digital narrative in the chaos of conflict can offer decisive advantages. Growing research has helped advance our understanding of how conflict shapes social media, and conversely, how social media influences conflict dynamics. Social media reduces the cost of communication,[2] increases the speed of its dissemination, and provides passive polling of conflict actors.[3] In addition, data gleaned from social media can be used to study network processes,[4] public opinion,[5] political representation,[6] protests,[7] and a wide range of other phenomena.[8]

In this essay, we investigate how social media usage may be influenced by local conflict dynamics. We study Twitter usage by individuals based in Syria during the conflict, including data from 2014 to 2017. Instead of studying the content posted by individuals using Twitter, we focus on account activity as an indicator of changing offline dynamics. Narratives on social media may change not only because individuals change the type of content they post, but also because the composition of users posting from a certain location changes.

We compare the number of monthly accounts that were newly created as well as the number of newly inactive accounts in two areas with very different conflict dynamics: the rebel-dominated Jebel Saman district and the regime-controlled area of Latakia. Between 2014 and 2017, the Jebel Saman district, which includes the city of Aleppo, was subjected to heavy fighting, repeated changes in armed group presence, and was frequently the site of large government offenses against both civilians and opposition groups, all which resulted in high numbers of casualties. Jebel Sama's city Aleppo has been subject to some of the most intense fighting and changes in local control experienced throughout the period under investigation here. The Latakia district in the northwest of the country includes the city of Latakia and is a pro-government stronghold, and is also home to bases for the Russian Navy and Air Force, the latter of which started engaging in the civil war in October 2015. In both Latakia and the neighboring governorate of Tartus, the Alawi represent the majority of the local population. Since the beginning of the uprising, the Assad Regime has actively taken steps to promote sectarian divides, continuously emphasizing the regime's close link to the Alawi community.[9]

We find that changes in conflict dynamics, such as the end of the regime-led siege in Aleppo in December 2016, coincide with substantial changes in local account composition. When scholars and practitioners rely on geo-located social media posts to make sense of a conflict, taking into account changes in the local composition of active accounts is crucial, but rarely done.

Despite significant insights into the use of social media by civilians proceeding political conflict, few studies examine civilians' use of social media beyond initial mobilization. Research on wartime social media use by non-combatants has instead focused on issues such as awareness raising campaigns, especially when they draw international attention to a subnational issue.[10] Analyzing the strategy of doctor-activists, for example, Alasaad (2013) shows how this specific group used Facebook and YouTube to spread international awareness about a leishmaniasis outbreak in Deir Ezzor province in 2013.[11] In sum, the use of social media in the context of ongoing civil conflict, in particular when used by civilians, remains understudied and may be distorted by methodological issues, such as when neglecting changing patterns in users on the ground. In the following section we discuss social media usage during the Syria conflict and offer a descriptive analysis of activity patterns of Twitter users who identified their location in Syria between 2014 and 2017.

Social Media Activity During the Syrian Conflict

The Syrian conflict has been called the most socially mediated civil conflict in history,[12] with some commentators going so far as to claim that the internet itself has become a weapon of war.[13] Armed actors on all sides of the conflict use social media to communicate their (change in) allegiances, spread propaganda, and interact with both their domestic and foreign audience.[14] While major bans on social media platforms in Syria were lifted a few months prior to the outbreak of the conflict in 2011, Internet activity remains highly surveilled and controlled by the Syrian regime.[15] Countrywide Internet shutdowns have occurred numerous times, and the regime has strategically limited access to the Internet in certain governorates (administrative units equivalent to American states) as part of their broader repressive strategy.[16]

Because of the volume of social media content, and the relative inaccessibility of the country to researchers, Syria became a key testing ground for innovative methods for studying the conflict from afar. Experts warned that the abundance of social media data provided observers only the illusion of complete information about events in Syria.[17] But little research has been done on the effects of these often veiled limitations on or distortions of the data, or on the explicit ways in which dynamics of the conflict itself, such as changes in the composition of conflict parties and shifts in territorial control, directly impact the nature of social media discourse.

Measuring Account Activity

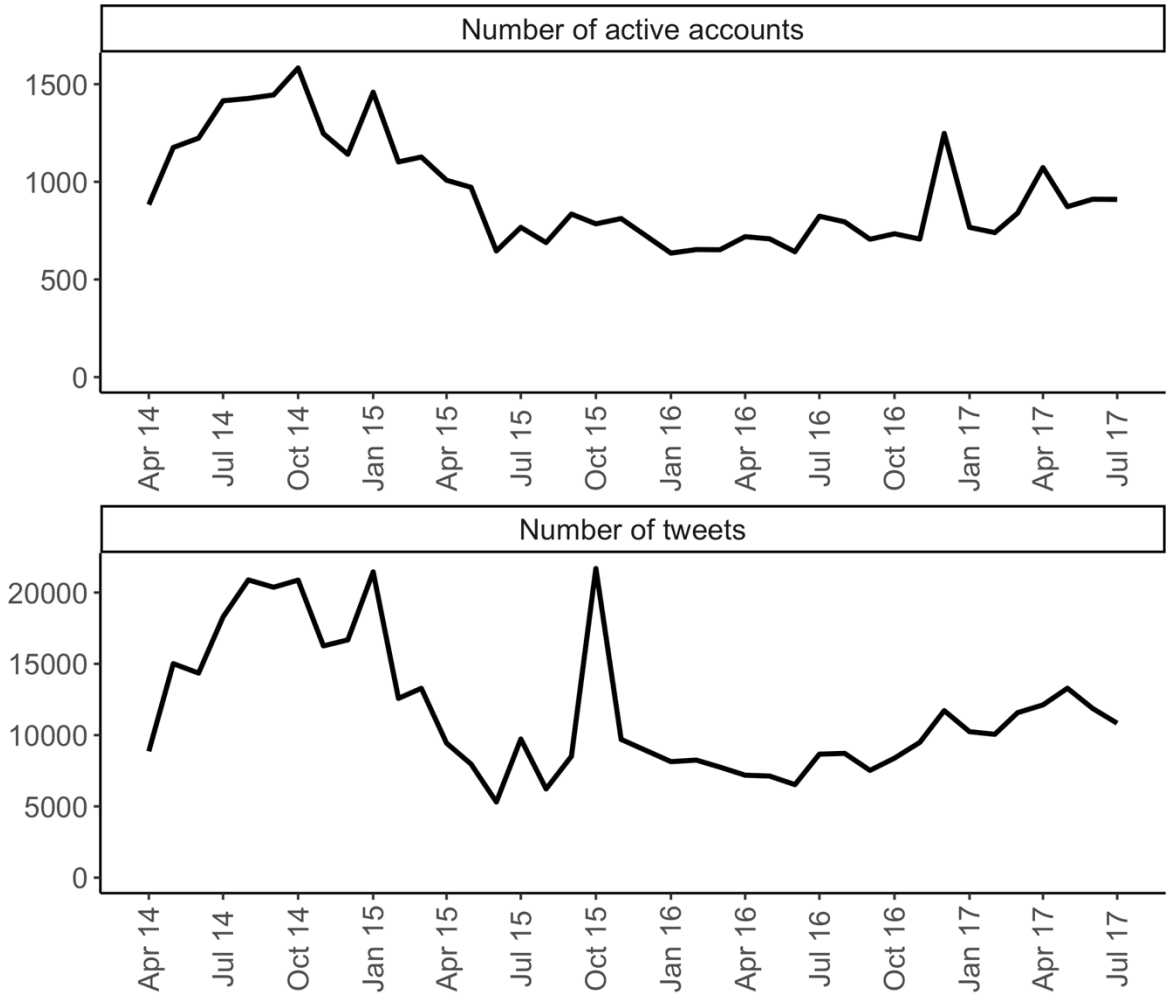
To understand how conflict affects social media behavior, we rely on geolocated Twitter posts. We collected the tweets in real-time, connecting to Twitter's POST statuses/filter endpoint to collect those that include longitude and latitude coordinates. Globally, 2-3% of tweets contain location coordinates, though approximately 12.91% of Arabic tweets are geotagged.[18] Since Twitter matches the parameters of a request up to a 1% ceiling, this process therefore provides between 7.7% ($\frac{1}{12.91}$) and 50% ($\frac{1}{2}$) of all tweets with coordinates.[19] We query the stored tweets for those that were sent from Syria from April 2014 until October 2017. This filter provides 474,223 tweets from 20,926 unique accounts.[20] The following indicators are calculated at the monthly level:

- Tweets: the number of tweets posted.
- Active accounts: the number of unique accounts.
- Account activity:
  - Accounts created: number of accounts created.
  - Accounts that go inactive: number of accounts that become inactive. An account is classified as inactive when it stops tweeting for at least three months. The start of inactivity is defined as the day on which they posted their last tweet.[21]

Figure 1 shows the number the number of active accounts (top panel) and tweets (lower panel) by month from April 2014 until July 2017.[22] The trends track each other before diverging. From April 2014 to June 2015, the number of active accounts and tweets in Syria follow the same n-shape: increasing through October 2014, the number of active accounts and tweets steadily declines into June 2015. In October 2015, however, there is a surge of tweets without a corresponding increase in accounts. The sharp increase in posting activity coincides with the start of the military intervention of Russian forces in the conflict. In December 2016, there is a surge of account creation without a corresponding increase in tweets, and this increase corresponds with the end of the siege in Aleppo. Other than these two spikes, the trends match: from lows around June 2016, both slowly increase during the remainder of the sample.

To further investigate the trends in Figure 1, Figure 2 shows account activity in two districts of Syria, Jebel Saman and Latakia. In November 2016, regime forces had circled the remaining densely populated rebel-held areas in the Eastern part of Aleppo, and, in a coalition with the Russian Air Force, submitted the area to intense bombardment for

Figure 1: Activity by accounts geo-located in Syria, April 2014 - July 2017.



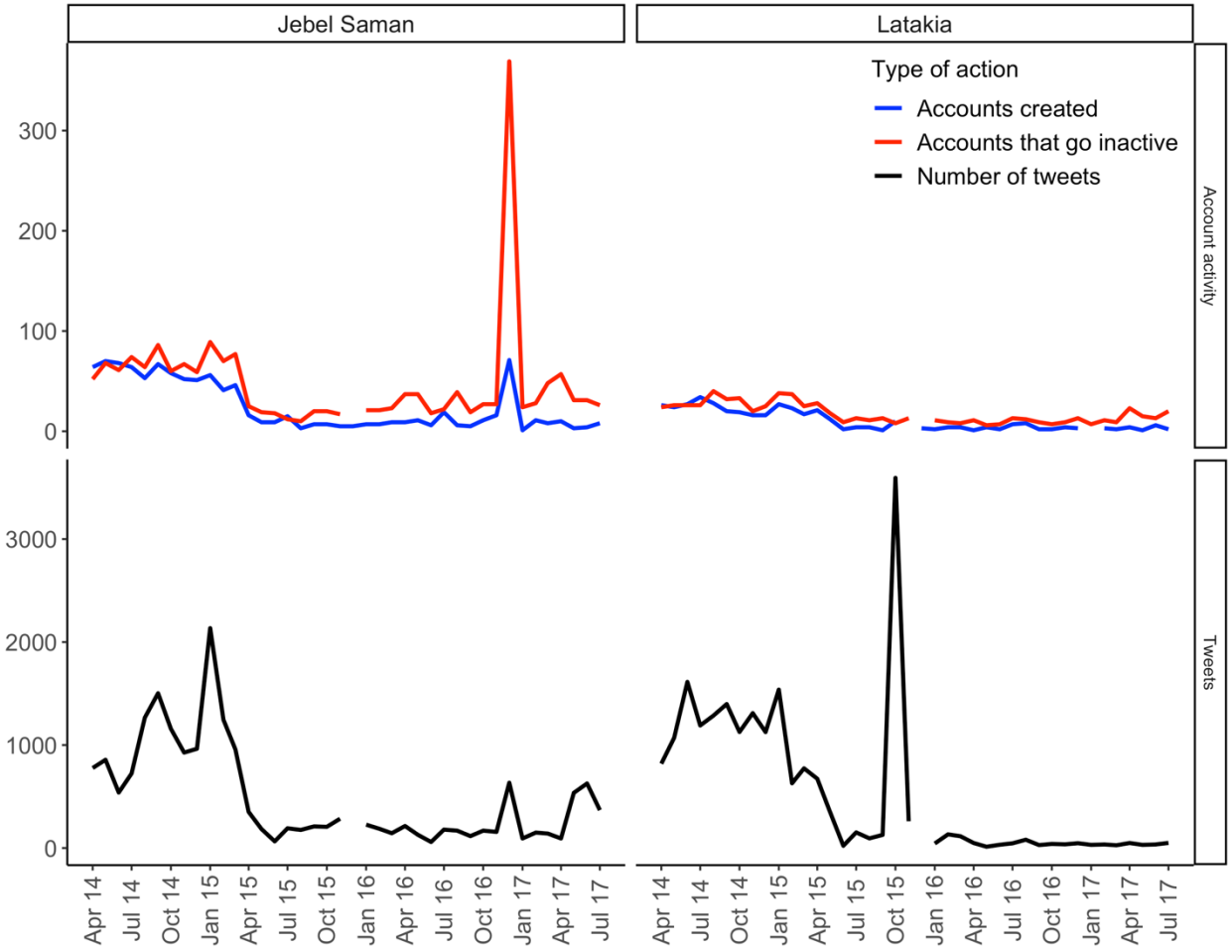
Note: The top panel shows the number of active accounts per month. The lower panel shows the number of tweets per month. After June 2015, they track each other much less closely than before. Counts are at the country level.

twelve days, targeting core civilian infrastructure such as hospitals.[23] On December 13, a highly complex ceasefire was negotiated which involved the surrender of weapons and transfer of all remaining rebels to other territories, resulting in an estimated relocation of one hundred thousand individuals.[24] The relocation continued through December 15. The siege left the city destitute with tens of thousands dead, and many more close to starvation.

Figure 2 shows activity in Jebel Saman (left panels) and Latakia (right panels). The top panels show the monthly number of created accounts as well as the number of monthly accounts that become inactive. The lower panels show the monthly number tweets of during the same time period. These four panels show that district level trends sometimes match national ones and other times do not, and the times of divergence coincide with important offline events.

Before June 2015 the district-level trends are comparable to the national trends. In both districts, the n-shaped pattern reappears, including the steady drop through June. Jebel

Figure 2: Changes in geo-located account activity, April 2014 - July 2017



Note: The top panel shows the number of accounts; the bottom shows the number of tweets. The left panel shows activity in Jebel Saman, an opposition stronghold, and Latakia, a pro-government district. The spikes in Jebel Saman correspond to the end of the Aleppo siege, and the spike in Latakia occurs the same month the Russian Air Force, based in Latakia, intervened in the civil war.

Saman's tweets increase after June 15, but not consistently. In Jebel Saman, the descriptive graphs reveal intermittent spikes in tweet activity but not the steady increase seen at the country-level. The graph shows that the end of the Aleppo siege was accompanied by a significant change in the local composition of Twitter accounts. In the sample studied here, 369 accounts become inactive, and 71 new ones are created. The large increase in inactive accounts in December 2016 starkly contrasts with country trends, which actually show an increase in the number of active accounts.

An approximate estimate of the true number of accounts that go inactive is possible. Using the percentage of Arabic users who geotag their tweet (12.91%) from Huang and Carley 2019 and assuming Twitter users from Syria geotag themselves at the same rate, then the true number of accounts that go active is approximately 2585 - $(\frac{1}{12.91} * 369 = 2858.25)$. The same calculation suggests that 549 $(\frac{1}{12.91} * 369 = 549.96)$ accounts activated at that time.

A closer look at the accounts that become inactive in Jebel Saman in December 2016 reveals similarities and differences to the overall sample of Syria-based accounts in the dataset. Qualitative studying of a random selection of account bios in both samples suggests similar types of accounts. In both the overall sample and the December 2016 Jebel Saman sample, only a small fraction of accounts tweet regularly (weekly). The accounts that go inactive in Jebel Saman have a higher average number of followers than the overall Syria sample, and post longer tweets (90 characters versus 68 characters). We also compare the sentiment of the tweets, counting the number of positive and negative words.[25] Both samples use almost the same number of negative words, but the sample of accounts that go inactive in Jebel Saman use significantly fewer positive words than the Syrian sample does. Fewer positive words despite longer tweets suggests a less positive sentiment within the group of accounts that become inactive compared to those that remain active. Overall, accounts that go inactive are more popular and less positively valenced than those that stay active.

Looking at the new accounts created and active in Jebel Saman in December 2016, we observe that more of the Twitter bios are in English than in the overall sample. New accounts are roughly as active as the average account in the overall sample. Similar to the accounts that went inactive that month in Jebel Saman, the newly active accounts tweet longer messages (93 versus 68 characters). New accounts use comparatively fewer positive *and* negative words, despite posting longer tweets. Overall, newly active accounts seem to therefore convey less sentiment than the overall sample of tweets for Syria.

In Latakia, the pattern of activity in 2015 and 2016 looks very different than in Jebel Saman. There is little to no variation in the number of accounts created or the number of accounts that cease activity. In contrast, posting activity rapidly decreased in the summer of 2015, only to then sharply increase in October of the same year. This rise in volume of tweets coincides with an overall spike in tweets geo-located in Syria in October 2015. While rises in Twitter activity in Jebel Saman - driven mainly by users based in Aleppo - are generally accompanied by a host of newly created accounts, changes in the number of tweets do not display this pattern in Latakia.

There are a number of possible explanations for these divergent patterns. Changes in posting frequency in Aleppo may be related to the death of individuals previously using Twitter, or to population movements within and outside of the area, either of which could contribute towards explaining the temporal variation in user composition. Alternatively, users in Aleppo may have chosen to abandon an old social media account and start a new one. Changes in account composition would then reflect changes in user behavior, rather than user composition. In light of the intense changes to the political and security context of the cases we studied, it is plausible to assume that the patterns of posting and account activity are a reflection of who was willing and able to be publicly active online. The indication of manipulated account activity at the outset of the Russian military involvement in the conflict further supports the argument that patterns of online activity are a function of changing conflict dynamics. We aim to further explore these dynamics in future work.

Finally, the rise in geo-located tweets in Latakia coincides with the overall activity visible in Figure 1 and the Russian Air Force's involvement in the conflict in October 2015, and a handful of accounts almost exclusively drive this activity. Moreover, these accounts appear to be manipulated. In our sample, 33 accounts tweeted from Latakia that month, but two authored 3,309 of all tweets; the third most active account only tweeted 156 times. These accounts present as a K-pop (Korean Pop Music) fan account, yet the second and third most active follow no accounts. In this paper's sample, 89.67% of the three accounts' tweets are from October 2015 and subsequent fan account activity is highly irregular - behavior consistent with manipulated accounts. These accounts still exist, i.e. they have survived Twitter removal of coordinated account networks, though their (infrequent) tweets still focus on K-pop.

Following the taxonomy in Leber and Abrahams 2021, these accounts appear to be inauthentic and ambiguous.26

The irregular but heavy tweeting suggests a bot, but the coordinated nature of the activity and the apolitical activity are more suggestive of a coordinated support group. Official foreign influence operations these accounts are not.[27] We cannot identify the actor or actors managing the accounts, though other studies of digital influence operations in the region means Russia or a member of the KUBE block are the most likely culprits.[28,29] Perhaps the most surprising feature of these accounts is that they do not appear to be part of a larger deception operation like often occurs in the region. This apparent isolation could be because manipulated accounts rarely geotag tweets, so those identified here could be part of a larger campaign invisible to this paper's sample.

Discussion

More than a decade after social media recorded and broadcast the first protests in the Middle East and North Africa, digital communication has become an everyday feature of modern conflicts. Much progress has been made in understanding the ways in which these tools are used for protest mobilization and coordination, yet beyond initial conflict onset, the everyday use of social media by civilians caught in the midst of war remains understudied.

While these results are descriptive, preliminary, and require further investigation, they emphasize that activity on social media reflects local changes in conflict dynamics. Shifts in territorial control at the local level may significantly change the composition of local users who are active on social media, resulting in shifting patterns of content that are attributable to shifting activity patterns, and not necessarily shifting sentiment of emotion of the same individuals. In future iterations, we will continue to build knowledge about how conflict dynamics affect social media usage.

Endnotes

[1] We thank Lucien Baumgartner for excellent research assistance.

[2] Andrew T Little, "Communication Technology and Protest," *Journal of Politics* 78, no. 1 (2015): 152–166.

[3] Thomas Zeitzoff, "How Social Media Is Changing Conflict," *Journal of Conflict Resolution* 61, no. 9 (2017): 1970–1991, doi:10.1177/0022002717721392.

[4] Daniel M Romero, Brendan Meeder, and Jon Kleinberg, "Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter," in *Proceedings of the 20th International Conference on World Wide Web* (ACM, 2011), 695–704, isbn: 9781450306324; Pablo Barberá et al., "The Critical Periphery in the Growth of Social Protests," *PloS ONE* 10, no. 11 (2015): 1–15, doi:10.7910/DVN/WCXK3Z.Funding, https://doi.org/10.1371/journal. pone.0143611.

[5] Thomas Zeitzoff, "Does Social Media Influence Conflict? Evidence from the 2012 Gaza Conflict," *Journal of Conflict Resolution* 62, no. 1 (2018): 29–63, doi:10.1177/0022002716650925; Nicholas Beauchamp, "Predicting and Interpolating State- Level Polls Using Twitter Textual Data 00 State-level public," *American Journal of Political Science* 61, no. 2 (2019): 490–503, doi:10.1111/ajps.

[6] Stan Oklobdzija, "Dark Parties: Citizens United, Independent-Expenditure Networks and the Evolution of Political Parties," in *Political Networks Workshops & Conference* (2018); Barberá, Pablo, Casas, Andreu, Nagler, Jonathan, Egan, Patrick J., Bonneau, Richard, Jost, John T., & Tucker, Joshua. (2019). Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data. *American Political Science Review, 113 no. 4 (2019)*, 883-901. doi:10.1017/S0003055419000352; Pablo Barberá and Thomas Zeitzoff, "The New Public Address System: Why Do World Leaders Adopt Social Media?," *International Studies Quarterly* 62, no. 1 (March 2018): 121–130, doi:10.1093/isq/sqx047.

[7] Marlon Mooijman et al., "Moralization in social networks and the emergence of violence during protests," *Nature Human Behaviour* 2, no. June (2018): 389–396, doi:10.1038/s41562- 018- 0353- 0, http://dx.doi.org/10.1038/ s41562-018-0353-0; Jennifer M Larson et al., "Social Networks and Protest Participation: Evidence from 130 Million Twitter Users," *American Journal of Political Science* 63, no. 3 (2019): 690–705, doi:10.1111/ajps.12436; Anton Sobolev et al., "News and Geolocated Social Media Accurately Measure Protest Size" (2020).

[8] Pablo Barbera and Zachary C Steinert-Threlkeld, "How to Use Social Media Data for Political Science Research," chap. 23 in *The SAGE Handbook of Research Methods in Political Science and International Relations* (London: SAGE Publications Ltd, 2020), 404–424.

[9] See e.g. Michael Kerr and Craig Larkin, *The Alawis of Syria: War, Faith, and Politics in the Levant* (New York: Oxford University Press, 2015).

[10] Abeer Najjar, "Othering the Self: Palestinians Narrating the War on Gaza in the Social Media," *Journal of Middle East Media* 6, no. 1 (2010): 1–30.

[11] Samer Alasaad, "War diseases revealed by the social media: massive leishmaniasis outbreak in the Syrian Spring," *Parasites & vectors* 6, no. 1 (2013): 1–3.

[12] Marc Lynch, Deen Freelon, and Sean Aday, "Blogs and Bullets III: Syria's Social Mediated War," *United States Institute of Peace, Peaceworks* 91 (2014): 5.

[13] Mohamed Hashem, "Q&A: In Syria the 'internet has become a weapon' of war," *Al-Jazeera*, 2015, http://www.aljazeera. com/indepth/features/2015/06/qa-syria-internet-weapon-war-150619215453906.html.

[14] Dana M Moss, "The ties that bind: Internet communication technologies, networked authoritarianism, and 'voice' in the Syrian diaspora," *Globalizations* 15, no. 2 (2018): 265–282, doi:10.1080/14747731.2016.1263079, https://doi.org/10.1080/14747731.2016.1263079.

[15] Freedom House, "Syria," *Freedom on the Net 2015*, 2015, https://freedomhouse.org/sites/default/files/resources/ FOTN%202015_Syria.pdf.

[16] Anita R Gohdes. "Repression Technology: Internet Accessibility and State Violence." American Journal of Political Science, 64:3 (2020):s 488-503. https://doi.org/10.1111/ajps.12509

[17] Lynch, Freelon, and Aday, "Blogs and Bullets III: Syria's Social Mediated War"; Megan Price, Anita Gohdes, and Patrick Ball, "Documents of war: Understanding the Syrian conflict," *Significance* 12, no. 2 (2015): 14–19, doi:10.1111/j.1740-9713.2015.00811.x.

[18] Binxuan Huang and Kathleen M. Carley. "A Large-Scale Empirical Study of Geotagging Behavior on Twitter". *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Network Analysis and Mining.* 2019: 365-373.

[19] For more information on working with Twitter, see Zachary C. Steinert-Threlkeld (2018). *Twitter as Data* (Elements in Quantitative and Computational Methods for the Social Sciences). Cambridge: Cambridge University Press.

[20] The data collection process failed to collect tweets for December 2015.

[21] We choose to measure account inactivity retroactively through the date of the last tweet because actually deleting an account can pose difficulty, in particular when accessing Twitter from a mobile device or through an app. We assume that users who either switch accounts or stop using their account will more likely stop actively using it long before they eventually close it (if at all). Users may sign up for a new account, and access this through the Twitter app without actually deleting their old account simultaneously. In addition, Twitter's API only confirms that an account no longer exists but not its date of cessation.

[22] We restrict the figures to three months before the end of our observation period to allow for the classification of inactive accounts.

[23] Annabelle Böttcher, "News Analysis Humanitarian Aid and the Battle of Aleppo," 105, no. 1 (2017): 2-3.

[24] Laila Bassam, Angus McDowall, and Stephanie Nebehay, *Battle of Aleppo ends after years of bloodshed with rebel with- drawal*, 2016, https://www.reuters.com/article/us- mideast- crisis- syria/battle- of- aleppo- ends- after- years- of- bloodshed-with-rebel-withdrawal-idUSKBN1420H5; Atlantic Council, "Breaking Aleppo," *The Atlantic Council of the United States*, 2017, 1–70.

[25] We split the corpus by language and apply separate, translated lexicons to each using the R package *syuzhet*. Saif M Mohammad and Peter D Turney, "Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon," in *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, CAAGET '10 (USA: Association for Computational Linguistics, 2010), 26–34. We use the get_nrc_sentiment function for English and Turkish Tweets. For Arabic tweets we replicate the procedure using the Arabic translation of the NRC Emotion Lexicon, created by Mohammad Salameh, Saif M. Mohammad, and Svetlana Kiritchenko http://saifmohammad.com/WebDocs/Arabic%20Lexicons/nrc_emotion_ar.txt

[26] Andrew Leber and Alexei Abrahams. Social media manipulation in the MENA: Inauthenticity, Inequality, and Insecurity. *POMEPS Special Series* 2021.

[27] Alexandra A Siegel. Official Foreign Influence Operations: International Broadcasters in the Arab Online Sphere. *POMEPS Studies 43.* (2021).

[28] Akin Unver and Ahmet Kurnaz. Russian Digital Influence Operations in Turkey 2015-2020. *POMEPS Studies 43.* (2021).

[29] Marc Owen Jones. Tracking Adversaries: The Evolution of Deception and Manipulation Tactics on Gulf Twitter. *POMEPS Studies 43.* (2021).

The Project on Middle East Political Science

The Project on Middle East Political Science (POMEPS) is a collaborative network that aims to increase the impact of political scientists specializing in the study of the Middle East in the public sphere and in the academic community. POMEPS, directed by Marc Lynch, is based at the Institute for Middle East Studies at the George Washington University and is supported by Carnegie Corporation of New York and the Henry Luce Foundation. For more information, see http://www.pomeps.org.

COVER PHOTO: SMEX NEWSLETTER AND THE ILLUSTRATOR NOUGAT.