# Slanted Narratives, Social Media, and Foreign Influence in Libya

Shelby Grossman

Stanford Internet Observatory, USA


Katie Jonsson

Stanford Internet Observatory, USA


Nicholas Lyon

Stanford University, USA


Lydia Sizer

Tufts University, USA .

Foreign social media outlets with political interests can use social media platforms to attempt to influence a country's politics. In fragile contexts such as Libya, where both social media penetration and information uncertainty are high, these efforts could be especially impactful. In this study, we use Libya as a case to assess how social media content about a polarizing, conflict-related event varies by the country of the information producer. We created a dataset of Facebook posts about a strongman's attack on Tripoli in 2019 (N=16,662). We find that more than half of the posts originated from outside Libya and that posts from countries aligned with the former Tripoli-based government are biased towards it and posts from countries aligned with the eastern-based strongman are biased toward his forces. However, many Pages are not slanted: the correlations are instead driven by a smaller number of hyperpartisan Pages. Our findings have implications for our understanding of how social media content – especially from abroad – could shape citizen perceptions of the legitimacy of competing political actors.

Shelby Grossman: shelbygrossman@stanford.edu; Katie Jonsson: kjonsson@stanford.edu; Nicholas Lyon; njlyon@stanford.edu; Lydia Sizer: Lydia.Sizer@tufts.edu.
Date submitted: 2022-04-25.

## Introduction

People are increasingly relying on social media for news. Thirty-six percent of American adults younger than 30 report often getting news from social media (Shearer, 2018). In Argentina, another country for which there is reliable data on this topic, 78% of people who follow the news use social media as a news source (Newman, 2020). This reliance is potentially worrisome given the lack of transparency on many social media accounts: anyone in any country can create a Facebook Page or Twitter account and claim to be a credible news outlet. Social media users may not know who created the content on their feeds and may trust accounts that resemble news providers. This is troubling – especially in uncertain, conflict-riven contexts – as foreign social media content could be part of an overt or covert influence operation.

To what extent does social media content about a country originate outside of that country, and does foreign content differ from domestic content? We assess these questions by investigating the content and origin of public posts on Facebook Pages about Libya in April 2019. During that month, eastern-based strongman Khalifa Haftar and his self-styled Libyan Arab Armed Forces (LAAF, also known as the Libyan National Army (LNA)) attempted to seize Tripoli, the capital, from the UN-backed Government of National Accord (GNA) headed by then Prime Minister Fayez al-Sarraj. The military battle for Tripoli played out as an internationalized civil war, with Egypt, France, Russia, Saudi Arabia, and the UAE engaging in unconventional warfare in support of the LAAF, and Turkey and Qatar supporting the GNA through similar tactics.

Using a Facebook-owned product called CrowdTangle, we created a dataset of Facebook Page posts, in Arabic and English, about the start of Haftar's Tripoli offensive in April 2019. In the month that followed Haftar's offensive there was a flood of Facebook posts describing the event from media outlets of varying legitimacy. Some framed the story as an internationally recognized government defending itself from a warlord, while others framed it as a general

bringing security to Libya.[1]  For each post, we manually coded the locations of the Page administrators using Facebook's Page Transparency feature: Facebook's best guess at the administrators' locations, making an assessment based on all of the data at their disposal. The data are useful particularly in cases where administrators attempt to use their self-declared location to conceal their real location. We then compiled a dictionary of pro-LAAF terms and one of pro-GNA terms to create a *GNA slant* measure for all posts and looked at the relationship between slant – the percent of words in a post aligned with the GNA or LAAF – and Page administrator location. In total, our dataset includes 16,662 posts (89% in Arabic, 11% in English).[2]

We find that more than half of the posts about the Tripoli offensive were from Pages with a plurality of administrators based outside of Libya, and 10% of posts have a plurality of administrators who have opted into hiding their location. We also find that there is a substantively significant relationship between the location of the content producers[3] and the slant of the post: posts from countries aligned with the GNA (Turkey and Qatar) are more pro-GNA and posts from countries aligned with the LAAF (the UAE and Egypt) are more pro-LAAF. Yet many Pages are not slanted; these correlations are instead driven by a subset of highly biased Pages. To better understand this surprising finding, we dig into the Pages producing the most slanted coverage and find that an overwhelming majority on both sides have ties to foreign countries. Overall, these findings suggest that during this important month Facebook content about Libya was heavily influenced by foreign actors and that the war in Libya is fought fiercely in non-kinetic domains by state and non-state actors.

Foreign social media influence operations can be especially destabilizing in places where foreign countries meddle in domestic politics – countries like Syria, Sudan, Venezuela, Mozambique, and Libya. Since the early 2000s, social media has evolved from a tool used by weaker actors in the international system to set agendas and compel stronger actors to change behavior, to a tool used by stronger states to manipulate regional and international dynamics

---

[1]Haftar calls himself a general.

[2]We focus on posts from Facebook Pages only because 1) CrowdTangle does not provide data on posts from individual profiles, and 2) CrowdTangle does not provide administrator location information for Groups.

[3]We will use the terms administrator and content producer interchangeably.

without overt use of force on the cheap (Robinson, 2013). Information warfare on social media can be especially attractive to states due to the plausible deniability if the operative is uncovered and the ability to make content appear as if it originated locally. Dozens of social media platform public takedowns of state-backed disinformation operations show that these information operations are prolific and target countries where other forms of foreign meddling proliferate (Brooking et al., 2020; Twitter, 2021). We note, however, that many of the foreign social media posts about Libya are not part of an influence operation.

Libya is a useful case for several reasons. First, there is evidence that Egypt, Russia, Saudi Arabia, and the UAE have conducted online disinformation campaigns about Libya on Twitter and Facebook.[4] Second, at the time the research was conducted, both al-Sarraj and Haftar struggled for legitimacy, and we can see how online media lowers the barriers for each side to attempt to delegitimize the other (Zhuravskaya et al., 2020). Third, while researchers and reporters pay much attention to how foreign actors use social media to influence citizens in high-profile cases such as the US, Hong Kong, and the UK (Bail et al., 2020), there has been less attention to the vulnerabilities of developing countries. Fourth, Libya is a useful case because of new developments to its media space: social media flourished after Muammar Qadhafi fell. In 2019 31% of Libyans said their primary source of information was Facebook (Wee and Li, 2019). And fifth, research suggests that information operations can be more effective when people are less certain about the truth (DellaVigna and Gentzkow, 2010); because conflict creates high degrees of information uncertainty, Libya shows the potential consequences of these campaigns.

This study contributes to three literatures. First, we contribute to research on social media and polarization. By conducting a deep dive into social media content around a highly salient and polarizing conflict, where the content producers were both domestic and foreign, we can shed light on why social media may have such effects. We show that the content of social media posts is itself polarized, indicating that content producers can frame the same events in dissimilar and slanted ways. Second, Christopher Walker and Jessica Ludwig have introduced the concept of "sharp power" to describe authoritarian information warfare.[5] Whereas soft power aims to win "hearts and minds" (Nye Jr, 2008), sharp power is used to manipulate peoples' information environment. This concept emerged to capture an increasingly important

---

[4]Russia: DiResta and Grossman (2019), Egypt/UAE/Saudi Arabia: Kassab and Carvin (2019).
[5]See Walker and Ludwig (2017) and Walker (2018).

feature of influence campaigns, but we believe we are the first to attempt to measure the potential scale of sharp power operations in a particular context. Third, there is an important literature on consumption of false stories (e.g. Guess et al., 2018). Increasingly, however, disinformation – defined as the intentional creation and sharing of information with the intent to deceive – takes the form of unfalsifiable hyperpartisan content (King et al., 2017), spread by fake accounts. In this paper we show that the foreign backers of competing political actors leverage unfalsifiable content to shape perceptions of the legitimacy of the opposing side.

### Clashing narratives: A theory of slanted social media in conflict settings

There is a large literature suggesting social media and online media can politically polarize citizens (Sunstein, 2018; Bakshy et al., 2015; Allcott et al., 2019) – though perhaps no more than offline media (Gentzkow and Shapiro, 2011) – and shape political attitudes (Guess, 2020; Walter, 2019) and behavior, including vote choice (Mitts, 2019) and protest (Enikolopov et al., 2020). The forces behind the production of politically polarized media, however, have received less attention. Who is creating the polarizing content and what motivates them to create it? In a domain that seems to collapse physical space, what is the role of geopolitics in the creation of polarizing content? In this section we first introduce a theory for why foreign actors will find social media information operations appealing in countries experiencing civil war. Second, we theorize how the location of the social media content producer relates to the slant of the content *about* countries experiencing civil war.

### *Why foreign actors run information operations on social media*

Countries in the midst of civil war are some of the most important contexts for understanding social media and political polarization, as civil war creates high levels of information uncertainty. Access to traditional media may be disrupted,[6] increasing reliance on social media for information. Individuals in these contexts may be more persuadable; research shows individuals will be more likely to update their beliefs in response to new information when they are more uncertain (DellaVigna and Gentzkow, 2010). Supplying information in a highly uncertain environment (e.g. in the middle of conflict) is one way of influencing citizen beliefs about the status of the conflict. Moreover, a large literature shows that media can shape individuals' allegiances

---

[6]See for example Williams (2004).

and behaviors. Yanagizawa-Drott (2014) finds exposure to propaganda on radio broadcasts increased participation in the Rwandan genocide; Adena et al (2015) show that mass media affected support for the Nazi party; and Dellavigna et al (2014) show that nationalistic media heightens inter-ethnic animosity.

There are several incentives for factions involved in civil unrest, and especially their external patrons, to exploit these tools, especially social media. First, social media gives a variety of actors, including foreign governments, domestic political actors, domestic citizens, and international bodies, insight into public opinion about a policy or individual (Zeitzoff, 2017). Social media is particularly likely to serve this role during civil war, when there are greater obstacles to running public opinion surveys. This can incentivize foreign actors to manipulate the perception of public opinion, for example by engaging in astroturfing. This refers to a deceptive campaign to inflate the appearance of support for an issue or individual. The term astroturfing comes from the notion of deception: the support is not *grass*roots, but rather akin to astroturf. These campaigns make it appear as if accounts are ordinary citizens speaking on their own volition (Keller et al., 2020). By leveraging the social influence of perceived peers, astroturfing can be used in attempts to shift allegiances in conflicts, much as peer influence has been shown to work in other domains (Turner, 1991). Empirically, many known disinformation campaigns – for example, networks suspended and publicized by Facebook and Twitter – have engaged in astroturfing, including in operations trying to sway public opinion about the popularity of particular political actors during civil war.[7]

This raises a second attractive aspect of social media for foreign actors: plausible deniability. Foreign governments can hide behind affiliated non-state social media accounts, which themselves can take steps to hide their Internet locations. This allows foreign actors to interfere in corrosive ways that advance their interests, while continuing to publicly sign onto popular peace initiatives.

Third, foreign actors can use social media as a tool of interference or intervention in civil wars to augment diplomatic tools. Foreign actors can *overtly* use social media in

---

[7]See for example a report on a network Twitter suspended that originated in Egypt, Saudi Arabia, and the UAE (Grossman et al., 2020), and a report on a network Facebook suspended that originated in Russia (Diresta and Grossman, 2019).

the same way that Jones and Mattiacci have theorized rebels use social media (2019): to spread positive narratives about their own interventions in the conflict, and negative messaging about foreign actors supporting an opposing side. Diplomatic tools are essential to preventing, mitigating, and even ending civil wars; however by substituting social media for true diplomatic engagement, foreign actors distort the purpose of this engagement, which is traditionally aimed at decreasing the amount of asymmetric information about intentions between warring factions (Regan and Aydin, 2006). Although using social media in this way may prolong civil war, foreign actors may have enough physical distance from the conflict that they have an incentive to manipulate the conflict in their favor using social media rather than support events to mitigate the violence.

Finally, in the last generation of civil wars, foreign actors involved have recognized that the center of gravity in such irregular warfare – what the factions are fighting over – is not territory *per se*, but the "hearts and minds" of the people (Smith, 2006). The intent of battles is less to defeat opposing factions militarily, but to win over a population's support while often fighting among them (Cohen and Horvath, 2006). In this context, foreign actors may be incentivized to employ social media to manipulate people's perspectives about a conflict without having to use force in ways that might alienate them.

For these reasons, foreign actors interested in winning "hearts and minds" in a country experiencing civil war will be tempted to run overt or covert information operations on social media. Winning hearts and minds matters as "civilian attitudes affect civilian actions"; civilians who support a rebellion, for example, will be less likely to share intelligence about rebel groups with the government (Mikulaschek et al., 2020, p. 775). Because media can affect individual allegiances and civil wars are, in part, a battle for legitimacy, belligerents and their allies should invest in slanted media to support their positions.

### *Social media slant and location of content producer*

Slant can be measured in two different ways. First, it can be measured by assessing *which* events media outlets cover (Baum and Zhukov, 2015). Second, slant can be measured by *how* a topic is covered. This measure is important, because often the target of information is a domestic audience, where certain events – for example battles or scandals – are impossible

to ignore. Slant of coverage about a particular event or issue is a common outcome in many political and economic studies of the media (e.g. Gentzkow and Shapiro, 2010) and the outcome of interest for this paper.

Theories of traditional media strategy typically assume that media outlets have financial incentives and that there are costs to crafting and sharing stories (Napoli, 1997). These assumptions may not always hold for social media. How should these theories be adjusted (1) when the cost of distributing stories is close to zero, as is the case on social platforms (Zeitzoff, 2017), and (2) in contexts where foreign state-tied outlets have a large presence in the social media space and there are non-financial incentives to production? Because this paper focuses on the Middle East and North Africa, and the relevant actors are primarily (though not exclusively) a set of countries in the region, we focus on media and authoritarianism as most countries in the region are not democracies (Fish, 2002).

We expect content originating from countries with authoritarian regimes will generally align with the foreign policy positions of the governments of those countries. This should be the case for several reasons. First, we expect state media to create content that is aligned with the objectives of that government. This content can be globally influentatial; Metzger and Siegel show that in one period of time RT (Russian state media) was the most shared news source on Twitter, both in Arabic and English, for content about Syria (2019). But state media is not the only way that governments try to control the online political narrative. Governments may run disinformation campaigns, covertly hiring digital marketing firms to push their narrative (DiResta et al., 2019). In these instances of disinformation, we again expect that content originating in a given country will align with the political objectives of the country's government.

Second, we expect that in countries with authoritarian regimes, even media outlets not tied to the state may publish government-aligned content. These outlets may fear reprimand or censorship for publishing non-state aligned stories. Less directly, reporters may be motivated by future career prospects (Baron, 2006) and, in countries where the public sector is a large employer, they may self-censor or slant stories to avoid alienating future employers. Budget constraints may further push news outlets to rely on government sources for stories (Herman and Chomsky, 2010).

Third, we expect consumer demand to play a weaker role on social media and thus content may more likely be supply-driven and government-aligned. Studies of traditional media look at the interaction of financially motivated media outlets and consumer demand, expecting that outlets bias their stories to align with the perceived opinions of their readers to increase their audience (Gentzkow and Shapiro, 2006). While some social accounts are profit-motivated – directing users to websites with ads, for example – covertly or overtly government-tied media outlets may not be financially motivated but rather policy motivated.

An exciting literature is emerging that uses rigorous empirical strategies to assess the impact of foreign overt propaganda campaigns (e.g. Bail et al., 2020; Chapman and Gerber, 2019). In this study, we are attempting to answer a question that is prior in the causal chain: how much of the content that people are consuming *overall* comes from foreign actors, and how does the content vary by location of content producer? Critically, we want to capture grassroots content, propaganda, and disinformation. For future work that aims to build on Chapman and Gerber (2019) and assess the effect of covert and overt information operations on public opinion, we believe our study will provide an important stepping stone: documenting the portion of the information environment that includes content from abroad, and how that information varies from domestically produced content.

**Foreign interference in Libya and the fight for political legitimacy**

***Evolution of foreign interference in Libyan affairs since 2011***

The competition for foreign influence in Libya has increased since the 2011 NATO military intervention that contributed to the ouster of former dictator Muammar Qadhafi. Following the revolution, the foreign actors that backed the rebels, including France, Qatar, the UK, and the US, sought stronger diplomatic, commercial, and security relationships with new, post-Qadhafi governments.

Libya's transitional governing institutions faltered quickly because of inexperience with democracy; inadequate post-conflict support from foreign partners; the Qadhafi legacy of corruption and bureaucratic incompetence; and destabilizing internal conflict and terrorist threats (Chivvis et al., 2012). In this context, many Western governments and private sector entities tempered their enthusiasm about engagement, and post-revolution governments became less

picky about the foreign support they received and regional powers sought to exploit increasing factionalism in Libyan politics and society to advance their own, often ideological, foreign policy agendas. Various Islamist politicians and militias received support from countries like Turkey and Qatar, while anti-Islamist factions received support from the UAE, Egypt, France, and, later, Russia (Fishman, 2017). With the eruption of a second civil war in 2014, foreign patrons shifted their focus from influencing the course of democratic transition in Libya to directly fueling the violent conflict between warring factions. By 2019, foreign support for rival factions almost single-handedly sustained the conflict in the capital, Tripoli, between anti-Islamist leader Khalifa Haftar, based in the east, and a diverse mix of opponents from Libya's west – a conflict that has displaced hundreds of thousands of people and resulted in the deaths of hundreds of civilians (Badi, 2019). While this paper focuses on 2019, we note that in October 2020 the GNA and Haftar signed a peace deal. On March 15, 2021, the Haftar-aligned House of Representatives gave its shaky vote of confidence to the new interim government, the Government of National Unity, at a historic parliamentary session. Relations between this new government and the House of Representatives collapsed over the course of 2021, and, as of 2022, there are still ripe opportunities for foreign influence campaigns in Libya.

### *The battle for political legitimacy*

Libya's second civil war was a battle not only for territorial control, but also – just as importantly – for political legitimacy (Gluck, 2015; International Republic Institute, 2019). The Libyan people have not approved a post-revolution constitution that could confer clear legitimacy on a government, largely because no Libyan government since 2011 has held a constitutional referendum. Additionally, no government has provided the adequate basic services and security necessary to earn popular legitimacy. Since the formal division of political institutions in 2014, rivals defined political legitimacy differently to empower themselves at the expense of their opponents. For example, the Government of National Accord (GNA) in Tripoli emphasized international recognition in its definition of legitimacy because UN Security Council resolutions recognized it as the legitimate Libyan government.[8] On the other hand, Haftar's rival government in the east defined legitimacy in terms of parliamentary and militia power, in part because Libya's internationally recognized and democratically elected parliament, the House of Representatives, had been located in the eastern city of Tobruk since 2014. Further-

---

[8]UN Security Council resolution 2259, adopted December 23, 2015, 7598th meeting.

more, Haftar branded his loosely grouped militia force, the LAAF, as a professional military (even though his force did not demonstrate Huntingtonian characteristics of professionalism) (Huntington, 1981).

In an effort to re-unify the country following civil unrest in 2014-2015, UN facilitators tried to construct power-sharing between rival factions by giving them power within different segments of a post-conflict, unity government. The anti-Islamist-dominated, Tobruk-based House of Representatives would be the internationally recognized parliament; a new, consultative High State Council would be dominated by many Islamist-leaning politicians; and the GNA would serve as an overarching, representative executive branch. This agreement was never fully realized. Between 2014 and 2021 members of the House of Representatives who did not support anti-Islamist Haftar either voluntarily boycotted the Tobruk-based parliament or faced persecution, including forced disappearance by pro-Haftar militias in the east (Zaptia, 2019a; Amnesty International, 2019). The House of Representatives as an associated, unrecognized government served as a political ally to Haftar, whose ambitions to control Libya fueled conflict as LAAF forces slowly conquered territory in parts of the country. Haftar and his political allies and foreign patrons refused to confer legitimacy on the GNA because they sought to politically dominate Libya themselves. They also saw the GNA as beholden to Islamist groups such as the Muslim Brotherhood, which pro-Haftar elements, such as the UAE, designate as a terrorist organization, in part to erode the political legitimacy of Islamist groups in Libya (Watanabe, 2016; Zaptia, 2019b).

Social media has served as an important forum for debate over political legitimacy in Libya, in part because public faith in traditional media in Libya is very low. During the revolution, Libyans relied more heavily on foreign satellite television news broadcasts than on local media sources that the Qadhafi dictatorship controlled (Mekay, 2011). The political and security vacuum led to systematic attacks on the media from unaccountable militias – including threats against journalists made on Facebook – resulting in the deterioration of the quality of local news and rampant self-censorship (Freedom House, 2016).

### *The evolution of Facebook's role in Libya*

In this context, Libyans both inside and outside the country have turned to social media to arbitrate conflict. Protestors used Facebook as an important organizing tool during the 2011 revolution despite regime attempts to block online communications (Ali and Fahmy, 2013). It remains the most popular social media platform in the country, especially for Arabic-langage communication (Global Voices, 2018). In the absence of reliable local reporting on the unfolding conflict, citizens turned to Facebook for news about where clashes were breaking out and what disruptions they could expect (Global Voices, 2018). Militias used Facebook to spread propaganda and fake news about their opponents; organize armed coalitions and recruit new members; find and sometimes kill individual opponents; and even buy and sell weapons (Walsh and Zway, 2018; Gatehouse, 2016).

The growing post-revolution unrest, culminating with Haftar's offensive in Tripoli in April 2019, provided an opening for foreign actors to shape Libyan affairs through social media. But as foreign actors became more invested in the outcome of Libya's post-revolution conflict, these actors – aligned with either the GNA or the LAAF (see Table 1) – plunged deeply into this largely Libyan debate on social media, via Facebook and other outlets, during Haftar's ongoing assault. As Haftar failed to take over Tripoli as quickly as he had anticipated, he and his foreign patrons, especially the UAE and Saudi Arabia, considered all pillars of power available to shift the balance in his favor: military, financial, and – importantly – rhetoric about political legitimacy. While Haftar's patrons poured military equipment, advisors, and funding into his kinetic campaign against Tripoli, they also augmented the eastern government's campaign to erode the GNA's remaining legitimacy through information warfare on social media (Facebook, 2019). Therefore, this study analyzes foreign involvement in shaping political-actor legitimacy during the 2019-2020 Tripoli offensive because of the prevalence of available evidence that explains how foreign actors weaponized social media in Libya – and how they may try to do so in different contexts in the future.

### Data and approach

In this section we outline how we constructed the dataset and our methodological approach. The goal was to create a dataset of posts about a specific topic and the content producer location.

| GNA-aligned governments | LAAF-aligned governments |
|---|---|
| Turkey | UAE |
| Qatar | Egypt |
|  | Russia* |
|  | Saudi Arabia |
|  | France |

**Table 1:** Governments aligned with the GNA and LAAF in 2019. * – though we include Russia in the list of LAAF-aligned governments, we note that its foreign policy towards Libya was, at the time, more nuanced than this binary table suggests. We discuss these complications pertaining to Russia's alignment in greater detail in section 5 of this paper.

We created a measure that captures post slant and looked at the relationship between that slant and content producer location. To create this dataset, we focused on Haftar's offensive on Tripoli, which began on April 4, 2019: an event that dominated Libyan news for over a year and in which regional actors were invested in how Libyans perceived the offensive.

We used CrowdTangle, a social media monitoring platform owned by Facebook, to identify public Page posts about the start of the offensive. Note that we do not include posts from private user accounts, which CrowdTangle does not provide access to, nor Groups, which CrowdTangle provides access to but lack locational information. We are focusing on public Pages – a Facebook feature used by brands, celebrities, and media outlets to connect with people. Popular media tends to suggest that Groups are often "echo chambers" for hyperpartisan content, suggesting that if anything focusing on Pages may understate the degree of highly slanted content overall.

For our dataset, we set the start date at March 30, 2019, as Haftar's regional backers may have been aware of the offensive ahead of time, and the end date at April 30, 2019. This range captures the period when the offensive was highly salient. We used 70 search terms (listed in Table 5 of the Appendix), split between Arabic and English words and phrases, selecting words and phrases to capture posts about the Tripoli offensive. For example, we used the search terms "Operation Flood of Dignity Tripoli" in English and Arabic. This references

Haftar's military offensive to capture western Libya and Tripoli. The final dataset included 14,910 unique posts in Arabic and 1,752 unique posts in English.[9]

Many social media studies have struggled with what is termed "the denominator problem" (Tufekci, 2014): it can be difficult to understand whether a social media dataset is representative of what an individual may be exposed to. We believe there is currently no better method for solving this denominator problem for Facebook data than the approach just described, though 1) the ideal denominator would show all online content, not just Facebook Page content, and 2) the CrowdTangle dataset has limitations. At the time of data collection Crowd-Tangle included 99.89% of Pages with more than 100,000 followers, though it has progressively less coverage as Pages have fewer followers.[10] Their coverage for smaller Pages is dependent on regular CrowdTangle users like us adding Pages. When any CrowdTangle user adds a Page, it is there for everyone. It is difficult to say how this may bias the dataset, as the people who use CrowdTangle – journalists, academic researchers, brands, Facebook investigators – use it for different purposes. In short, our findings should be interpreted as more representative of larger Pages and less representative of smaller Pages. We note, however, that our dataset does include some small Pages: 4 percent of Pages in the dataset have less than 1,500 followers, 12 percent of Pages in the dataset have less than 10,000 followers, and 53 percent of Pages in the dataset have less than 200,000 followers.[11]

We read each post to confirm that it was about the Tripoli offensive and deleted any that were irrelevant. Next, we visited the post's Page to record information from the Page Transparency feature, which provides Facebook's best guess about Page administrator location: a useful tool if the guess differs from the user's self-declared location. Facebook uses all of the data at its disposal to make these location assessments. For example, Figure 1 shows Page Transparency information for the Al Jazeera English Page (Al Jazeera is based in Qatar).

In addition to listing specific countries, Page Transparency sometimes includes locations

---

[9]We dropped a small number of posts that were in Italian.

[10]The Crowdtangle coverage estimate for the time of data collection is here: https://web.archive.org/web/20210120170600/help.crowdtangle.com/en/articles/1140930-what-data-is-crowdtangle-tracking.

[11]Of the 1,487 unique Facebook Pages in our dataset, the number of Page Likes for 202 Pages was not exported by CrowdTangle. The 5 percent and 14 percent figures are reported for the set of Pages with non-missing Page Like data.
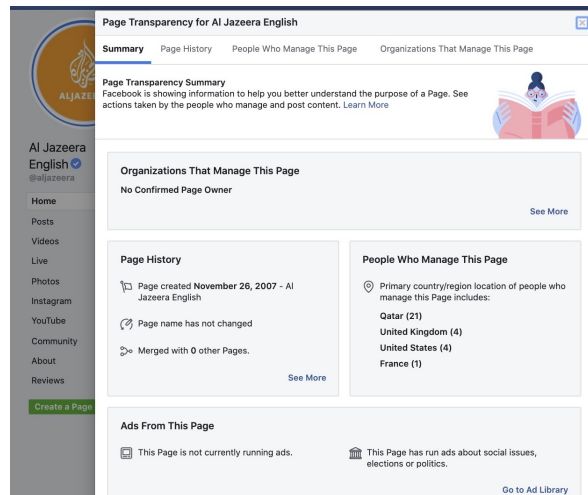
**Figure 1.** Page Transparency information for Al Jazeera English

as (for example) "Location hidden (1)" or "Not available (2)." Less often, Page Transparency shows no information at all. Though we lack a crystal clear understanding of the "Location hidden" category, it may still be an important one. Facebook's website says users can submit a form to request that Facebook hide their location – for example, for safety reasons.[12] But it also says that prior to October 2019 certain Pages could switch their location to hidden without a form. As such, it appears that Page administrators could opt into the location hidden status, making it a theoretically important category for our purposes. In our dataset, 1,596 posts – or 9.6% – are classified as originating from an administrator with "Location hidden," making the analysis of this category a matter of practical as well as theoretical importance.[13] "Not available" indicates that Facebook cannot confidently determine an administrator's location. Because of our uncertainty about the reason for that determination, we do not think of it as a meaningful category. Facebook will sometimes not show any information about Page administrators for Pages with small audiences.[14] Some Pages have a tie for the country location

---

[12]facebook.com/help/323314944866264

[13]As seen in Figure 6 of the Appendix, the proportion of posts coming from "Location hidden" Pages does not increase over the period under study, suggesting that new Pages are not being created with their location hidden to mask online activities towards Libya.

[14]We collected the Page Transparency information in December 2019 and January 2020. In some unusual cases, location information changes over time, presumably if Page administrators travel. Our

of a plurality of their administrators; in these cases, we break the tie at random and assign the post to a single country.

We believe we are among the first political scientists to leverage this Page Transparency data for academic research. This is an important data source, as typically it is difficult for researchers to assess where social media content producers are based. On Twitter, for example, just 1-2% of tweets are geolocated (Kruspe et al., 2021), and when users self-declare their location there are concerns about data validity as they may lie. Facebook has substantial data on user location and no obvious incentive to misreport in this product feature.

Next, we created two dictionaries, one with English and Arabic words and phrases that appeared in posts we deemed biased in favor of the LAAF and the other with words and phrases that appeared in posts we assessed favored the GNA. We created these dictionaries in 2020 in collaboration with Khadeja Ramali, an expert on social media in Libya. Examples of pro-LAAF terms we included are "liberation" (as in "Haftar will liberate Tripoli"), "Islamists" (as in "Haftar is fighting Islamists"), and "illegitimate" (as in "the GNA is illegitimate"). Pro-GNA words included "invaders," "renegade," and "war criminal"; these were words used to denigrate Haftar. The full dictionaries appear in the Appendix in Table 6. An example pro-LAAF post from the data is "General Haftar ordered the Lybian [sic] National Army to advance towards the positions of the Jihadist & Muslim Brotherhood militias inside Tripoli to pacify the capital and enable the government to protect civil society and organize elections. If this is serious, I hope the US, EU & UN support and mandate the LAAF to apply international and national law." An example pro-GNA post is "Warlord also understood to have private support of Saudi Arabia and the UAE" ("warlord" is a reference to Haftar).

To an outsider observer, some of our dictionary terms like "Qatar", "Libyan militias", or "Field Marshal" may appear neutral. Many of these terms, however, have special meaning to Libyan and regional audiences. These terms serve as coded language to inspire anger and fear in the audience, or to legitimize or delegitimize actors discussed in posts. In the case of the term "Qatar", content creators using this term typically sought to invoke an anti-Islamist fear of a spreading influence of political Islam, and the Muslim Brotherhood in particular, in Libya and the Middle East and North Africa region. Qatar was actively involved supporting

---

data captures the location information at a single point in time.

Islamist groups during the 2011 revolution; however Qatari involvement in Libya has waned drastically since 2012. Affiliating groups or actions with Qatar, however, especially during the nadir of relations between Qatar and its powerful, anti-Islamist neighbors in the Gulf in the 2010s, is meant to provoke anti-Islamist groups in Libya and abroad. Separately, identifying armed groups as "militias" was another subtle way for pro-GNA actors to delegitimize their enemies. There is a struggle for legitimacy among armed groups in Libya, many of whom try to portray themselves as official state security institutions despite the continued absence of unified police and military entities in the country. Being cast as "militias" undermines armed groups affiliated with different factions in the ongoing civil conflict in their efforts to shed this identifier and clothe themselves in the legitimacy of state institutions. Finally, posts that identify Haftar as "field marshal" versus "renegade" "warlord" serve a similar purpose – legitimizing Haftar's control over Libya's armed forces (which do not actually exist yet). A rump portion of the Libyan parliament promoted Haftar to field marshal in September 2016 to solidify his reputation in the eyes of Libyans and the international community as the leader of Libya's security apparatus.

### *Slant Measure*

We used information from the posts and the two dictionaries to create a measure of post slant that captures whether any given post favors the GNA or the LAAF. To construct that measure, we first conducted several processing steps to get the text of the CrowdTangle data into a workable format. First, we combined the Message and Description variables from the CrowdTangle data: the former shows the text of a Page's post while the latter shows a Description of any hyperlink that is included with the post. Because both types of content are accessible to viewers, we conduct our analysis on the combined text. Second, all Arabic posts were translated into English using Google Translate. This step mitigates the potential for our dictionary keywords to miss matches in the post data due to the use of different regional Arabic variants of the same word. For instance, our LAAF dictionary includes "Islamists" as key term. Depending on grammatical position, "Islamists" could be written as "islamiyyin," "islamiyyun," or several other variants. Because each of these variants is translated to the same "Islamists" equivalent in English by Google Translate, we are confident that our method picks up spelling differences or idiomatic variants across dialects. We did not use Google Translate on posts written in English. We then removed all hyperlinks and special characters in the text. Next, we split the

posts into unigrams and bigrams so that phrases in the text could be matched to terms from our two dictionaries. We stemmed all words in the unigrams and bigrams and then removed a set of 1,149 commonly used English stop words from the stop words dataset contained in the `tidytext` **R** package.

We also conducted these processing steps on our two dictionaries to ensure matches between the dictionaries and the text in the posts.[15] We then identified all pro-GNA terms by finding all the single-word matches between the pro-GNA dictionary and a post's set of unigrams and all two-word matches between the pro-GNA dictionary and a post's set of bigrams.[16] We limited this step to just unigrams and bigrams because – after the aforementioned text processing steps – none of our dictionary terms went beyond two words. We then did the same unigram and bigram matching for the pro-LAAF dictionary.

From the pro-GNA and pro-LAAF matches, we created a simple pro-GNA slant measure by subtracting the ratio of pro-LAAF term matches to total phrases (unigrams plus bigrams) in a post from the ratio of pro-GNA term matches to total phrases. Our pro-GNA slant measure for Facebook posts in the data, indexed by $i$ posts, is given by:

$$Slant_i = \frac{GNA_i - LAAF_i}{Unigrams_i + Bigrams_i}$$

$GNA_i$ is the number of times a term in the pro-GNA dictionary appears in post $i$ and $LAAF_i$ is the number of times a term in the pro-LAAF dictionary appears in post $i$. $Unigrams_i$ is the total number of unigrams in the post while $Bigrams_i$ is the total number of bigrams in the post. The resulting GNA slant measure is bound from -1 to 1, with numbers closer to -1 indicating more pro-LAAF terms compared to pro-GNA terms, and numbers closer to 1 indicating more pro-GNA terms compared to pro-LAAF terms. We normalize this measure for analysis. We include an example of all the steps involved in the creation of our slant metric in Appendix A.2.

---

[15]We did not split our dictionary terms into unigrams and bigrams, though, to avoid any multi-word dictionary terms being matched as unigrams with post text.

[16]For justification on the value of simply counting words in Arabic social media data, see: Siegel (2019).

In the analysis that follows, we look primarily at the relationship between the location of the plurality of administrators and the GNA slant measure, though we also show results for the raw number of pro-GNA and pro-LAAF terms.

To validate our slant metric, we had an independent coder identify the slant of a random sample of 500 posts from our dataset. The coding was conducted on a five category basis: (1) posts that are strongly pro-LAAF or strongly anti-GNA; (2) posts that are somewhat pro-LAAF or somewhat anti-GNA; (3) neutral posts without a clear slant; (4) posts that are somewhat pro-GNA or somewhat anti-LAAF; and (5) posts that are strongly pro-GNA or strongly anti-LAAF. This manual coding was done using text in the language of the original Facebook post.

To understand whether our slant measure captures meaningful variation in a reader's inferred slant, we break our slant metric into three meaningful categories: posts that are pro-LAAF (any post with a negative value on the raw slant measure); posts that are neutral (any post with a zero value on the raw slant measure); and posts that are pro-GNA (any post with a positive value on the raw slant measure). Figure 2 shows the mean hand-coded value and confidence intervals for each of these categories. Posts with a negative value on the raw slant measure are, on average, identified by the coder as being anti-GNA or pro-LAAF. Likewise, posts with a positive value on the raw slant measure are, on average, identified by the coder as being pro-GNA or anti-LAAF. Posts with a zero value on the raw slant measure are, on average, identified by the coder as being marginally pro-GNA. However, this estimate is much closer to neutral than the average hand-coding for posts our automated method identifies as pro-GNA (the difference between the two estimates is statistically significant at conventional levels). This gives us confidence that our slant measure does capture meaningful variation, on average, in the *relative degree of slant* between posts in this dataset.[17]

---

[17]During the validation exercise, the hand-coder identified 14 posts that had ambiguously partisan language. For instance, one post stated "He was accused of committing war crimes in Tripoli" without identifying who "he" was. This post can be found at https://www.facebook.com/maghreb-voices/posts/2411126519134437. These fourteen posts were removed from the validation exercise. Summary statistics for the raw slant measure at each hand-coded level are provided in Table 7 of the Appendix. Further validation checks indicate that the slant measure and the resulting analysis presented below are not sensitive to specific terms or phrases in the dictionaries (see Figure 7 of the Appendix).
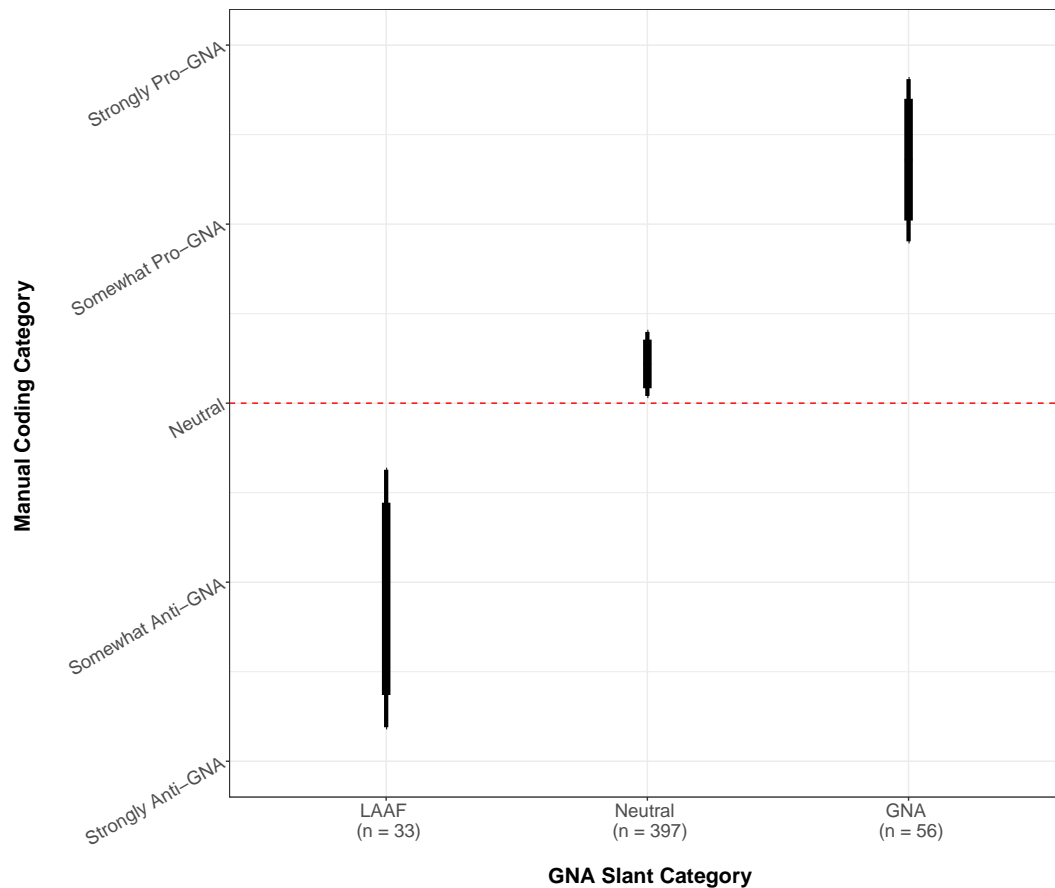
**Figure 2.** Validation Exercise. For this figure, the manual coding has been translated to numeric values, as follows: strongly anti-GNA = -2; somewhat anti-GNA = -1; neutral = 0; somewhat pro-GNA = 1; strongly pro-GNA = 2. Points show the mean hand-coded value for each of the three slant categories. The outer bars show 99 percent confidence intervals for this estimate while the inner bars show 95 percent confidence intervals. 14 posts that the hand-coder identified as being ambiguously slanted were removed from the analysis (see footnote 17).

The dataset on *slant* that we have created includes many relatively small Pages (less than 200,000 followers) that are (at least nominally) news sites. For example, the Page "Middle East Affairs," tied to middleeastaffairs.net, has 125,000 followers. These types of Pages, with often unclear ownership, are extremely prevalent and can dominate users' feeds, but academic researchers rarely study them. Many are likely covertly tied to governments. For example, in 2019 Facebook removed dozens of Pages targeting African countries, including Libya, that were linked to Yevgeny Prigozhin, a Russian oligarch deeply involved in unconventional warfare with ties to Russian President Vladimir Putin. The Pages present as news Pages with titles like Tripoli News Network (translated from Arabic) [18] and Libya News Network (translated from Arabic).[19] A Page that appears to be news may make users less skeptical, as we know that consumers assess the messages they receive based on what they know about sender credibility (DellaVigna and Gentzkow, 2010). Many of these Pages present as local news (like Tripoli News Network) and research (in the US) suggests that citizens are more likely to trust local news sources over national ones (Lakshmanan, 2018). Still, Libyan social media users are often savvy to foreign meddling in their online space.

### Does Facebook post content vary by location of the information producer?

This section assesses whether social media content about conflict events will align with the objectives of the government of the country where the content is produced.

There are 78 unique locations in our dataset for the plurality of Facebook Page administrators, including labels for "Location hidden" and administrator data "Not available." Figure 3 shows the geographic distribution of posts in the dataset. Pages with a plurality of administrators in Libya make up 48% of all posts – a plurality of content about the Tripoli offensive in April 2019 was produced within the country. The geographic distribution of posts is, in part, tied to a post's language: English language posts are more likely to originate from Pages with administrators in Libya, the US, the UK, and France, while Arabic language posts are more likely to originate from Pages with administrators in Libya, Egypt, Tunisia, Turkey, Jordan, and the Persian Gulf states.

---

[18]archive.fo

[19]archive.ph
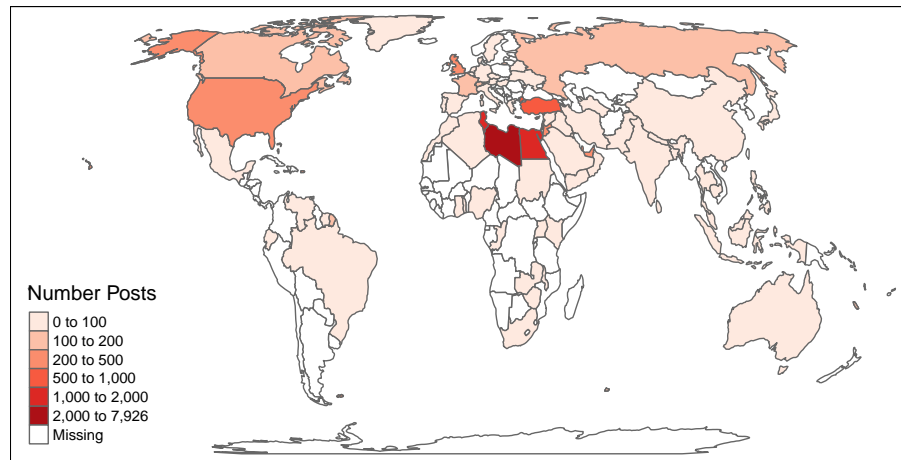
**Figure 3.** Number of posts by country.

The language used in a post is highly correlated with its partisan content. Posts written in English are, on average, 0.4 standard deviations more slanted in favor of the GNA than posts written in Arabic. This is true when estimates are produced using variation across all countries and when using within-country variation (see Table 8 in the Appendix).

Descriptive statistics for our measures of Page slant are included in Table 2. These measures are provided for the complete set of all posts as well as for ten relevant Page locations: Libya, the set of seven countries included in Table 1, Pages with their administrator location hidden, and the United States as a point of comparison. On average, posts about the Tripoli offensive during the month of April contain 0.22 words identified as pro-GNA keywords and 0.16 words identified as pro-LAAF keywords, resulting in a slight overall pro-GNA slant. The country-level statistics shown in Table 2 accord with how governments have aligned themselves toward the GNA and LAAF, respectively. Posts originating from Pages in Turkey and Qatar are slanted towards the GNA – with posts from those locations having, respectively, an average of 0.26 and 0.28 pro-GNA terms compared to an average of 0.05 and 0.08 pro-LAAF terms. Conversely, posts originating from Pages in Egypt and the UAE are slanted towards the LAAF, with averages of 0.34 and 0.26 pro-LAAF terms, respectively, and averages of 0.10 and 0.04 pro-GNA terms.[20]

---

[20]Figure 9 in Appendix A.4 shows the geographic distribution of our slant measure. Several of the

| Country | Number Posts | Mean GNA Keywords | St. Dev. GNA Keywords | Mean LAAF Keywords | St. Dev. LAAF Keywords | Mean GNA Slant | St. Dev GNA Slant |
|---|---|---|---|---|---|---|---|
| All posts | 16662 | 0.219 | 0.731 | 0.159 | 0.672 | 0.001 | 0.019 |
| Libya | 7926 | 0.269 | 0.864 | 0.138 | 0.580 | 0.003 | 0.018 |
| Egypt | 1685 | 0.100 | 0.507 | 0.338 | 1.052 | -0.004 | 0.020 |
| Location hidden | 1596 | 0.095 | 0.397 | 0.130 | 0.651 | -0.001 | 0.016 |
| Turkey | 790 | 0.258 | 0.679 | 0.052 | 0.264 | 0.007 | 0.023 |
| USA | 378 | 0.381 | 0.693 | 0.130 | 0.649 | 0.005 | 0.016 |
| Qatar | 298 | 0.282 | 0.813 | 0.084 | 0.323 | 0.004 | 0.015 |
| UAE | 266 | 0.038 | 0.243 | 0.256 | 0.707 | -0.006 | 0.020 |
| France | 113 | 0.310 | 0.642 | 0.115 | 0.458 | 0.003 | 0.018 |
| Russia | 105 | 0.076 | 0.359 | 0.105 | 0.390 | <0.001 | 0.011 |
| Saudi Arabia | 64 | 0.172 | 0.420 | 0.234 | 0.792 | <0.001 | 0.015 |

**Table 2:** Descriptive statistics are included for all posts and select countries. Statistics for the pro-GNA slant measure are provided using the raw data, as opposed to the normalized slant measure which is used in subsequent analyses.

In Table 3 we investigate whether the location of information production matters for post content in a regression framework. This table presents the results from 30 separate regressions: three different outcomes – normalized GNA slant, LAAF keyword count, and GNA keyword count – regressed separately on ten different country indicators of substantive interest. We run separate regressions for each country indicator – as opposed to a single regression with all location indicators – because we are interested in estimates of the slant of posts from each relevant country compared to all other posts. Results from regressions with all location indicators support the same conclusions (see Table 9 in the appendix). All regressions include language as a fixed effect and cluster standard errors at the Facebook Page level. For all coun-

---

countries that appear to be the most slanted – such as Australia and Zimbabwe – are those with very few Facebook posts about the Tripoli offensive in April 2019, but the minimal content posted from these places was very slanted.

tries except France, Russia, and the United States, there is a large and significant relationship between country of origin and post content. Estimates for the Libya, Qatar, and Turkey indicators in the standardized slant model are positive – indicating pro-GNA slant – and statistically significant at conventional levels. Estimates for the Egypt, Qatar, and UAE indicators are all negative – indicating pro-LAAF bias – and statistically significant at the 0.05 level. Because the outcome is standardized, estimates should be interpreted in terms of standard deviations: for instance, posts from Libyan Facebook Pages are estimated to be 0.19 standard deviations more slanted towards the GNA than non-Libyan posts. Though the negative estimate on the Saudi Arabia indicator is not statistically significant at conventional levels for the standardized slant outcome, posts from Saudi Facebook Pages are estimated to have fewer GNA keywords than non-Saudi posts (significant at 0.05 level).

Turkey and the United Arab Emirates are the countries that are estimated to be the most biased towards the GNA and the LAAF, respectively. Posts originating in Turkey are estimated to be 0.29 standard deviations more slanted *towards the GNA* than all other posts and posts originating in the United Arab Emirates are estimated to be 0.39 standard deviations more slanted *towards the LAAF* than all other posts. The substantive significance of these estimates is exemplified by the following set of three posts – (1) a neutral post with a standardized slant measure close to zero; (2) a pro-GNA post with a standardized slant measure close to 0.29; and (3) a pro-LAAF post with a standardized slant close to -0.39. Pro-GNA phrases are bolded and pro-LAAF phrases are underlined in the two excerpts, below.

1. "Places of control and deployment of the Government of National Accord forces and the movements of the forces loyal to Haftar in the battle of #Tripoli" (with an embedded video): *standardized slant of -0.079; raw slant of zero.*[21]

2. "...The Italian prime minister expressed regret that Tripoli was under an undue attack that brought the country back to the atmosphere of war after it was at the door of the solution to its crisis, and Mr. Conte confirmed Italy's decisive rejection of this

---

[21]A Facebook post from the Al Jazeera Channel, 24 April 2019. Translated from original Arabic. Available at https://www.facebook.com/aljazeerachannel/posts/10157673652374893. Note that this neutral post has a slightly negative standardized slant value even though its raw slant value is zero; the standardized values should be interpreted in relative rather than absolute terms.

| | Standardized Slant | LAAF Count | GNA Count |
|---|---|---|---|
| Libya | 0.19*** | −0.06* | 0.13*** |
| | (0.04) | (0.03) | (0.03) |
| Egypt | −0.29*** | 0.19** | −0.11*** |
| | (0.06) | (0.08) | (0.03) |
| France | −0.11 | < 0.01 | −0.05 |
| | (0.11) | (0.07) | (0.08) |
| Qatar | 0.15** | −0.08** | 0.07 |
| | (0.05) | (0.02) | (0.09) |
| Russia | −0.17 | −0.04 | −0.19 |
| | (0.15) | (0.04) | (0.10) |
| Saudi Arabia | −0.17 | 0.10 | −0.13** |
| | (0.09) | (0.13) | (0.05) |
| Turkey | 0.29** | −0.11*** | 0.04 |
| | (0.11) | (0.02) | (0.06) |
| UAE | −0.39** | 0.10 | −0.18*** |
| | (0.15) | (0.08) | (0.03) |
| USA | −0.04 | 0.03 | − < 0.01 |
| | (0.07) | (0.04) | (0.05) |
| Location Hidden | −0.13* | −0.04 | −0.11*** |
| | (0.06) | (0.03) | (0.02) |

$^{***}p < 0.01,\ ^{**}p < 0.05,\ ^{*}p < 0.1$

**Table 3:** Coefficients from separate regressions of each outcome – *Standardized GNA Slant*, *LAAF Count*, and *GNA Count*– on a single country indicator are reported above (the above table reflects 30 individual regressions). These regressions control for language $\in$ (Arabic, English) fixed effects. Standard errors clustered at the Facebook page level are included in parentheses. Each regression is estimated on a set of 16,662 observations across 1,487 unique Facebook pages.

**destabilizing** attack...": *standardized slant of 0.295; raw slant of 0.007*, signifying a post slanted towards the GNA.[22]

3. "...It is worth noting that the General Command of the Libyan Arab Armed Forces, led by Field Marshal Khalifa Haftar, announced on 4 April the launch of the 'Flood of Dignity' operation to eliminate <u>extremist</u> terrorist groups in Tripoli.": *standardized slant of -0.408; raw slant of -0.006*, signifying a post slanted towards the LAAF.[23]

 

 

The first post is informative without introducing normative language – markers of media that would be considered relatively neutral. The second post demonstrates language that is slanted towards the GNA: the post makes it clear that the Italian government rejected Haftar's "destabilizing" attack on Tripoli, thereby legitimating the GNA – indeed, this post comes from a GNA media account. The third post exemplifies a post that is slanted towards the LAAF by associating the invasion of Tripoli with the fight against "extremist terrorist groups." Haftar has previously accused the GNA in Tripoli of working with terrorists and the LAAF has sought international legitimacy, in part, by making claims of fighting against ISIS in Libya. These examples demonstrate that our slant metric captures meaningful variation in bias towards the GNA or the LAAF, respectively, and that the magnitudes of the coefficients reported in Table 3 represent meaningful differences in the partisan content of Facebook posts.

These findings from the standardized slant model in Table 3 largely hold for the count of LAAF and GNA keywords: post locations associated with content slanted towards the GNA – Turkey, Qatar, and Libya – either post content with more GNA keywords or post content with fewer LAAF keywords. The reverse is true for post locations associated with content slanted more towards LAAF – Egypt, Saudi Arabia, and the UAE. Posts from Pages with a plurality of hidden administrators are, on average, more slanted towards the LAAF and are less likely to post GNA keywords. Posts originating from the United States are not associated with slant or keyword usage in favor of either the LAAF or the GNA. Posts associated with

---

[22]Excerpted from a Facebook post from The Media Office of the President of the Presidential Council of the Government of National Accord, 8 April 2019. Translated from original Arabic. Available at https://www.facebook.com/GNAMedia/posts/2311860269071355.

[23]Excerpted from a Facebook post from the Libyan News Agency, 16 April 2019. Translated from original Arabic. Available at https://www.facebook.com/lananewspage/posts/2076120459173844.

France and Russia, respectively, are estimated to be slanted in favor of the LAAF but these associations are not estimated precisely.

Consistent with the data, Russia's foreign policy posture toward Libya during the time Russia-based administrators created the posts in this study suggests that such posts would be more neutral than others. Unlike Turkey and the UAE, countries that took a clear, partisan side in the Libyan conflict, the Russian government took a more bifurcated approach. Militarily, Russia has long provided support for Haftar and the LAAF, in part because the government sees Haftar as a vehicle for fulfilling lucrative, Qadhafi-era military contracts Russia had worth approximately U.S. $2 billion (VOA, 2010). But diplomatically, the Russian foreign ministry has sought to maintain relationships with all major factions involved in the civil conflict (Ramani, 2020). Maintaining a more neutral public stance in its social media posts would be consistent with Russia's overt, diplomatic strategy in Libya.

The fact that posts from the U.S. are similarly neutral on average could be due to many factors. Unlike other countries analyzed, it is likely that Facebook posts originating in the U.S. are less influenced by U.S. government positions on the Libyan conflict due to America's press freedom. Additionally, the Donald Trump administration sent mixed signals as to its position on the conflict, particularly in April 2019, the month of the social media content analyzed here (Kirkpatrick, 2019).

The slanted nature of posts from major, regional players supports the view that especially in countries like Libya, which are not strategically significant to great powers like the US, there is space for major, regional powers like the UAE and Turkey to use the information space to support their allies and proxies in the local conflict. During the period under study in this paper, Russia adopted a slightly different method for advancing its foreign policy interests in Libya. It is a method that it has applied in other parts of Africa, as well as in the United States in 2016: manipulating elections (Akinola and Ogunnubi, 2021). One Russian "sociologist" and his translator were jailed in Libya on charges of interfering in Libyan politics – specifically, supporting the former dictator's son, Saif al-Islam Qadhafi, in his political rehabilitation efforts ahead of expected elections (Higgins and Walsh, 2020).

Though the coefficient on Saudi Arabia is consistently anti-GNA, it is not statistically

significant across models. We suspect this may be a function of sample size: in our dataset only 64 posts originated on Pages where a plurality of administrators were in Saudi Arabia.

### Is a majority of content slanted toward the position of the government where the content was produced?

In this section we assess whether a majority of social media content about conflict events produced in autocratic settings is slanted towards the government's position, given incentives to conform in authoritarian regimes. Figure 12 in Appendix A.8 visually displays density plots of the post-level slant data for each of the ten countries included in Table 3, above. These plots show that, across all countries, posts are clustered around zero slant, indicating that a large fraction of posts are not explicitly slanted towards either the LAAF or the GNA. Rather, a subset of posts in each country appears to drive partisan content.

To assess whether most Pages, or only a few Pages, are driving country-level slant estimates, we zoom in on Page-level data for Turkey and the UAE – the most GNA-slanted and LAAF-slanted countries in the dataset, respectively. Figure 4 shows density plots for all Facebook Pages in Turkey and the UAE, respectively, that posted at least 20 times about the Tripoli offensive during April 2019. These plots suggest a high degree of Page-level variation in partisan content within countries. Several Turkish and UAE Pages did not post any slanted content during this period. This pattern holds for Qatar, Egypt, and Libya as well. To better understand the minority of partisan Facebook Pages, we analyze who is behind them in the following section.
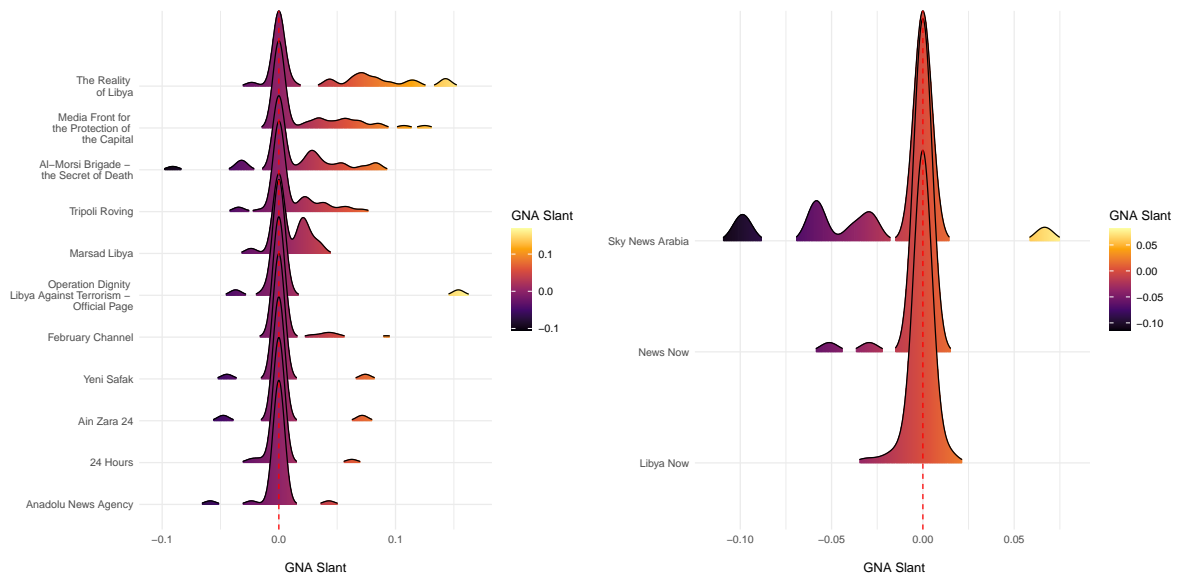
**Figure 4.** Page-level slant distributions for Turkey and the UAE. Left panel: Distribution of the GNA slant measure for the set of nine Facebook Pages with a plurality of Turkish administrators that posted at least 20 times during April 2019. Right panel: Distribution of the GNA slant measure for the set of three Facebook Pages with a plurality of Emirati administrators that posted at least 20 times during April 2019.

### Who is behind highly slanted Pages?

To better understand the surprising finding that a minority of Pages were creating the slanted content, we created a Page-level dataset, subsetted it to Pages with at least 20 posts, and identified (1) the top 10 most pro-LAAF Pages and (2) the top 10 most pro-GNA Pages, as defined by our dictionaries. For these 20 Pages, we attempted to understand their purpose and backers. We did this by consulting with a Libyan media analyst and using online open source investigation tactics, such as investigating the registration of affiliated websites. The results of this investigation are shown in Table 4 and Table 5.

There are two key takeaways. First, of the 10 most pro-LAAF Pages, only two are run by individuals in Libya. Rather, much of the pro-LAAF content appears to originate from Pages in countries with pro-LAAF associations: Egypt, Saudi Arabia, and the UAE. Second,

| Page name | Affiliated Countries | Notes |
|---|---|---|
| Abdul Rahim Ali | Egypt | Page for the individual who is the editor in chief of Albawabh News (see below) and a member of the Egyptian Parliament. Content about Libya on this Page - which often links to Albawabh News articles - condemns Turkish and Qatari involvement and plays up the importance of Egyptian involvement. |
| Portal News (Albawabh News ) | Egypt | Two of the top pro-LAAF pages are associated with the Egyptian news site Albawabh News: (1) https://www.facebook.com/Albawaba.eg/ and (2) https://www.facebook.com/albawab-hnewspaper. These pages appear to post identical content. The editor in chief of Albawabh News is a member of the Egyptian Parliament (see above). Content about Libya published on this site frequently criticizes Turkish involvement in Libya and draws connections between Turkey, Qatar, the GNA, and the Muslim Brotherhood. |
| Arab ANHA | Kurdish region of Syria | This is the Hawar News Agency, a Kurdish news service tied to the Syrian Democratic Forces. Many of this Page's pro-LAAF posts have a strong anti-Turkish slant as well, which may explain why a Kurdish news service would be invested in following events in Libya. |
| Baghdad Post + | Saudi Arabia | Page for thebaghdadpost.com/ar; the Chairman is an Iraqi man who, according to some reports, has strong relations with the Ministry of Foreign Affairs in Saudi Arabia. Posts about Libya on this page emphasize the role of the LAAF in combatting terrorism and identify Qatari involvement in Libya as suspicious. |
| Libyan Satellite Channel | Libya | This Page is now down though it had administrators in Libya during the period of time captured in our data-set. Posts from this Page that are available in our data-set frequently reference Major General Ahmed Al-Mismari, the official spokesperson for the LAAF. Quotes attributed to Al-Mismari condemn the influence of Turkey and Qatar and emphasize the role of the LAAF in combating terrorism. |
| Middle East and North Africa Media Monitor | Austria Jordan | It is not clear whether any government-linked entity is behind this Page. Current Page administrators are based in Austria and Jordan, but administrators were located in Austria, Egypt, and France for the period covered in our data-set. Articles on the linked website paint Turkish involvement in Libya in a negative light, with accusations of the plunder of Libyan wealth through an association with the Muslim Brotherhood. |
| Rose Al-Youssef News Portal - Official Page | Egypt | The *About* section for this Page states that it is affiliated with the Rose Al-Youssef Foundation, a journalism school in Cairo. The Page links to the online Rose Al-Youssef newspaper. Sources indicate that the paper is owned by the Egyptian state and its editor is selected by the Shura Council. Content about Libya on this Page covers Haftar's views and is critical of Turkish involvement. |
| Sky News Arabia | UAE | This Page is partly owned by Mansour bin Zayed Al Nahyan, a member of the ruling family and deputy prime minister in the UAE. Administrators for this Page are in seven different countries, though the majority are located in the UAE. Content about Libya on this Page references anti-Turkish and anti-GNA positions and frequently covers Haftar's views. |
| URGENT Libya Now | Libya | It is unclear who is behind this Page, though it has administrators in Libya. Content on this page frequently covers Haftar's views and rejects Turkish involvement in Libya. |

**Table 4:** The Ten Most Pro-LAAF Pages. One Page identified as pro-LAAF using our slant methodology is - upon inspection - pro-GNA. This misclassified Page is omitted from the table. Appendix A.11 discusses why this Page was misclassified.

| Page name | Affiliated Countries | Notes |
|---|---|---|
| Al-Morsi Brigade - The Secret of Death | Malta<br>Turkey | It is unclear who is behind this Page, though administrators are located in Malta and Turkey. The Malta connection is not surprising: Malta officially recognizes the GNA, has disrupted efforts to bankroll Haftar's campaign, and cooperates with Turkey in attempting to bring an end to the conflict. The *About* section of this Page references the "Armed forces of the February revolution," an allusion to the outbreak of the First Libyan Civil War in 2011 that deposed Muammar Qadhafi. Content about Libya on this page frequently calls Haftar a criminal. |
| Corner of the Heart of the Event | Libya | It is unclear who is behind this Page, though administrators are based in Libya. Content on this Page praises losses to Haftar's forces and celebrates collaboration between the GNA and Turkey. |
| Gharyan is Free | Jordan<br>Libya<br>UK<br>U.S. | Gharyan references a town that was Haftar's supply base for his Tripoli offensive. We are unable to determine who runs this Page, but we note that it has administrators in Libya, the US, Jordan, and the UK. Content is anti-Haftar, including a comparison between Haftar and Hitler, and highlights cooperation between the GNA and Turkey. |
| Libya News | Libya<br>Turkey | It is unclear who is behind this Page, though administrators are located in both Libya and Turkey. Content on this page frequently condemns Haftar and prominently links him to Russian mercenaries. Turkish and Qatari involvement is portrayed in a positive light. |
| Libya Odessa | Libya<br>Saudi Arabia | This Page is now down but - during the period in our data-set - had administrators in Libya and Saudi Arabia. Content from this Page in our data-set condemns Haftar - frequently referring to him as "rebel" - and voices support for "Volcano Anger Operation." |
| Media Front for the Protection of the Capital | Libya<br>Turkey | This Page is explicitly aligned with the GNA, as the *About* section says it is: "A media front that supports the Libyan army of the Government of National Accord against the rebel." |
| The New Arab | Qatar<br>Turkey<br>UK<br>U.S. | Owned by a Qatari media group and based in London, this Page has administrators in Turkey, the United Kingdom, and the United States. The news site was founded by Azmi Bishara, a former Israeli MP who is exiled in Qatar and has political connections to the Emir of Qatar. Content about Libya on this page includes references to "rogue Libyan general Khalifa Haftar", positive portrayals of Turkish involvement, and condemnation of Egypt's role. |
| Operation Volcano Anger | Libya | This Page is explicitly aligned with the GNA, as the *About* section says it is: "The official account of the media center for the Volcano Anger Operation." |
| The Reality of Libya | Libya<br>Turkey | It is unclear who is behind this Page. The *About* section references an address in Tripoli but most of the Page administrators are based in Turkey. Posts on this Page imply alignment with the GNA, for instance: "Strong clashes between *our forces* and Haftar militias in Tarhuna axis" (emphasis added). |
| Tribune Street Libyan | Libya<br>Turkey<br>UK | It is unclear who is behind this Page, though it has administrators in Libya, Turkey, and the UK. Posts on this Page imply alignment with the GNA, for instance: "Tripoli now, no, no, no, to the criminal Haftar...Yes to the army and police, yes to the civilian state." |

**Table 5:** The Ten Most Pro-GNA Pages

while more of the pro-GNA Pages appear to have direct connections to individuals in Libya, six of the top 10 Pages have associations with Turkey.

We also looked at the Pages that appeared most frequently in the dataset. The most prolific Page was facebook.com/libyaakhbar, with 429 posts (2.6% of posts in the dataset). The Page was on average slightly pro-LAAF. As of March 2020, the Page had just over a million followers. Interestingly, the Page's four administrators all have their location hidden, and its Twitter account, @libyaakhbar, is suspended.

Another prolific Page – with 281 posts – was Aljamahiria, the former state-run broadcast channel under Muammar Qadhafi. Recent research shows that a firm linked to Russian oligarch Yevgeny Prigozhin purchased half of this media outlet in 2019, and Facebook suspended the Page in December 2021 (Grossman et al., 2018). This is further evidence of the abundance of foreign-influenced Facebook Pages targeting Libya.[24]

## Conclusion

In this paper we have shown a correlation between the bias of Facebook posts about a salient event in Libya and the location of the Facebook Page administrators. In April 2019, Pages with a plurality of administrators in the UAE and Egypt were more biased in favor of the LAAF while Pages with administrators in Turkey and Qatar were more slanted toward the GNA. This aligns with the geopolitical interests of these countries. However, not all posts linked to these countries were slanted. We find that these correlations were driven by a smaller number of highly biased Pages. When investigating the most slanted Pages in the dataset, we found that an overwhelming majority of both pro-LAAF and pro-GNA Pages had strong links to foreign countries.

These findings suggest, first, that academics, analysts, and policymakers should interpret social media responses to events with deep skepticism. Many of these responses may originate abroad and social media accounts may be intentionally coy about where they are based. Moreover, this paper focused on Facebook Pages that were live as of early 2020, but at

---

[24]This discussion leads to questions about how social media users respond to slanted content, and where social media users who view this content live. Using a small sample of data, we explore these questions in Appendices A.12 and A.13.

the time of Haftar's Tripoli offensive in 2019 there were likely other Pages that have since been suspended. Since the start of 2019, there have been at least seven takedowns from Facebook and Twitter of foreign state-linked information operations that targeted Libya. This suggests that we are likely underestimating the extent to which foreign-originating content is flooding Libyan social media.

While we did not code the 16,662 posts for containing falsehoods, in the process of coding for other variables we observed that many slanted posts – possibly a majority – were not falsifiable. These often included hyperpartisan cheerleading posts like "Haftar will bring security to Libya"; this tone-setting content is in line with content researchers have seen the Chinese government push (King et al., 2017). This suggests that the current academic and policy focus on "fake news" (e.g. Guess et al., 2018) addresses only a subset of information operations, and that there may be a greater need for mechanisms to help citizens gain insight into the trustworthiness of social accounts, as opposed to specific posts. Facebook's Page Transparency is one step in that direction. These transparency tools may be especially useful during crisis events because of the magnitude of partisan influence from foreign sources.

Our findings suggest that the war in Libya from 2019-2020 was fought fiercely in non-kinetic domains by state and non-state actors. There is also a strong potential that conflict could reignite in Libya, creating a new opportunity for spillover into these non-kinetic domains. Information and ideas are just as important a battleground as territory in conflicts over legitimacy, including international interventions that result in regime change, revolutions, military coups, and counterrevolutions. The narratives pushed by foreign actors can shed light on their incentives. Further work is needed to generate causal identification strategies to assess the effectiveness of these social media information operations.

## References

Adena, M., Enikolopov, R., Petrova, M., Santarosa, V., and Zhuravskaya, E. (2015). Radio and the Rise of the Nazis in Prewar Germany. *The Quarterly Journal of Economics*, 130(4):1885–1939.

Akinola, A. E. and Ogunnubi, O. (2021). Russo-African relations and electoral democracy: Assessing the implications of Russia's renewed interest for Africa. *African Security Review*, pages 1–17.

Ali, S. R. and Fahmy, S. (2013). Gatekeeping and citizen journalism: The use of social media during the recent uprisings in Iran, Egypt, and Libya. *Media, War & Conflict*, 6(1):55–69.

Allcott, H., Braghieri, L., Eichmeyer, S., and Gentzkow, M. (2019). The welfare effects of social media. Technical report, National Bureau of Economic Research.

Amnesty International (November 1, 2019). Libya: No Contact with Abducted Politicians for Months: Siham Sergiwa.

Badi, E. (September 3, 2019). Libya's Hifter and the false narrative of authoritarian stability. *Middle East Institute*.

Bail, C. A., Guay, B., Maloney, E., Combs, A., Hillygus, D. S., Merhout, F., Freelon, D., and Volfovsky, A. (2020). Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017. *Proceedings of the National Academy of Sciences*, 117(1):243–250.

Bakshy, E., Messing, S., and Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239):1130–1132.

Baron, D. P. (2006). Persistent media bias. *Journal of Public Economics*, 90(1-2):1–36.

Baum, M. A. and Zhukov, Y. M. (2015). Filtering revolution: Reporting bias in international newspaper coverage of the Libyan civil war. *Journal of Peace Research*, 52(3):384–400.

Brooking, E. T., Kann, A., and Rizzuto, M. (2020). Dichotomies of disinformation: Project overview & codebook. https://github.com/DFRLab/Dichotomies-of-Disinformation.

Chapman, H. S. and Gerber, T. P. (2019). Opinion-Formation and Issue-Framing Effects of Russian News in Kyrgyzstan. *International Studies Quarterly*, 63(3):756–769.

Chivvis, C. S., Crane, K., Mandaville, P., and Martini, J. (2012). *Libya's Post-Qaddafi Transition: The Nation-Building Challenge*. Rand Corporation.

Cohen, E. and Horvath, J. (2006). Principles, imperatives, and paradoxes of counterinsurgency. *Military Review*, pages 49–53.

DellaVigna, S., Enikolopov, R., Mironova, V., Petrova, M., and Zhuravskaya, E. (2014). Cross-border media and nationalism: Evidence from Serbian radio in Croatia. *American Economic Journal: Applied Economics*, 6(3):103–32.

DellaVigna, S. and Gentzkow, M. (2010). Persuasion: Empirical evidence. *Annual Review of Economics*, 2(1):643–669.

DiResta, R. and Grossman, S. (2019). Potemkin Pages & Personas: Assessing GRU Online Operations, 2014-2019. , Stanford Internet Observatory.

Diresta, R. and Grossman, S. (2019). Potemkin pages & personas: Assessing GRU online operations, 2014-2019.

DiResta, R., Grossman, S., KH, and Miller, C. (2019). Analysis of Twitter Takedown of State-Backed Operation Attributed to Saudi Arabian Digital Marketing Firm Smaat.

Enikolopov, R., Makarin, A., and Petrova, M. (2020). Social media and protest participation: Evidence from Russia. *Econometrica*, 88(4):1479–1514.

Facebook (August 1, 2019). Removing coordinated inauthentic behavior in UAE, Egypt and Saudi Arabia.

Fish, M. S. (2002). Islam and authoritarianism. *World Politics*, 55:4.

Fishman, B. (July 12, 2017). The Qatar Crisis on the Mediterranean's Shores. *The Washington Institute for Near East Policy*.

Freedom House (2016). Freedom of the press 2016: Libya.

Gatehouse, G. (April 6, 2016). Weapons for sale on Facebook in Libya. *BBC*.

Gentzkow, M. and Shapiro, J. M. (2006). Media bias and reputation. *Journal of Political Economy*, 114(2):280–316.

Gentzkow, M. and Shapiro, J. M. (2010). What drives media slant? Evidence from US daily newspapers. *Econometrica*, 78(1):35–71.

Gentzkow, M. and Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4):1799–1839.

Global Voices (September 6, 2018). Netizen report: What role does Facebook play in Libya's civil war?

Gluck, J. (2015). 2. Constitution-building in a political vacuum: Libya and Yemen in 2014. *Annual Review of Constitution-Building Processes: 2014*, page 43.

Grossman, S., H, K., DiResta, R., Kheradpir, T., and Miller (2020). Blame it on Iran, Qatar, and Turkey: An Analysis of a Twitter and Facebook Operation Linked to Egypt, the UAE, and Saudi Arabia.

Grossman, S., Ramali, K., and DiResta, R. (March 20, 2018). Blurring the lines of media authenticity: Prigozhin-linked group funding Libyan broadcast media. *Stanford Internet Observatory.*

Guess, A. (2020). Does Exposure to Online Partisan Media Affect Political Attitudes and Behavior? .

Guess, A., Lyons, B., Montgomery, J. M., Nyhan, B., and Reifler, J. (2018). Fake news, Facebook ads, and misperceptions.

Herman, E. S. and Chomsky, N. (2010). *Manufacturing consent: The political economy of the mass media.* Random House.

Higgins, A. and Walsh, D. (June 18, 2020). How two Russians got caught up in Libya's war, now an action movie. *New York Times.*

Huntington, S. P. (1981). *The soldier and the state: The theory and politics of civil–military relations.* Harvard University Press.

International Republic Institute (May 15, 2019). New poll of 15 Libyan municipalities shows support for Councils, concerns over service delivery.

Jones, B. T. and Mattiacci, E. (2019). A manifesto, in 140 characters or fewer: Social media as a tool of rebel diplomacy. *British Journal of Political Science*, 49(2):739–761.

Kassab, M. and Carvin, A. (2019). A Twitter Hashtag Campaign in Libya: How Jingoism Went Viral. https://medium.com/dfrlab/a-twitter-hashtag-campaign-in-libya-part-1-how-jingoism-went-viral-43d3812e8d3f.

Keller, F. B., Schoch, D., Stier, S., and Yang, J. (2020). Political astroturfing on twitter: How to coordinate a disinformation campaign. *Political Communication*, 37(2):256–280.

King, G., Pan, J., and Roberts, M. E. (2017). How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *American Political Science Review*, 111(3):484–501.

Kirkpatrick, D. D. (April 19, 2019). Trump endorses an aspiring Libyan strongman, reversing policy. *New York Times*.

Kruspe, A., Häberle, M., Hoffmann, E. J., Rode-Hasinger, S., Abdulahhad, K., and Zhu, X. X. (2021). Changes in Twitter geolocations: Insights and suggestions for future usage. *arXiv preprint arXiv:2108.12251*.

Lakshmanan, I. A. (August 22, 2018). Finally some good news: Trust in news is up, especially for local media. *Poynter*.

Mekay, E. (February 24, 2011). One battle over Libya is fought via the media; Qaddafi has news outlets that are well financed, but opponents are savvy. *The International Herald Tribune*.

Metzger, M. M. and Siegel, A. A. (2019). When State-Sponsored Media Goes Viral: Russia's Use of RT to Shape Global Discourse on Syria.

Mikulaschek, C., Pant, S., and Tesfaye, B. (2020). Winning Hearts and Minds in Civil Wars: Governance, Leadership Change, and Support for Violent Groups in Iraq. *American Journal of Political Science*, 64(4):773–790.

Mitts, T. (2019). From isolation to radicalization: anti-Muslim hostility and support for ISIS in the West. *American Political Science Review*, 113(1):173–194.

Napoli, P. M. (1997). A principal-agent approach to the study of media organizations: Toward a theory of the media firm. *Political Communication*, 14(2):207–219.

Newman, N. (2020). Digital news report. *Reuters Institute for the Study of Journalism, University of Oxford.*

Nye Jr, J. S. (2008). Public diplomacy and soft power. *The annals of the American academy of political and social science*, 616(1):94–109.

Ramani, S. (April 7, 2020). Russia's strategy in Libya. *RUSI.*

Regan, P. M. and Aydin, A. (2006). Diplomacy and other forms of intervention in civil wars. *Journal of Conflict Resolution*, 50(5):736–756.

Robinson, P. (2013). Media as a driving force in international politics: The CNN effect and related debates. *E-International Relations*, 17.

Shearer, E. (2018). Social media outpaces print newspapers in the U.S. as a news source. *Pew Research Center.*

Siegel, A. A. (2019). Using social media data to study Arab politics. *APSA MENA Politics Section: MENA Politics Newsletter*, 2(2):33–36.

Smith, R. (2006). *The utility of force: The art of war in the modern world.* Penguin UK.

Sunstein, C. R. (2018). *#Republic: Divided democracy in the age of social media.* Princeton University Press.

Tufekci, Z. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8.

Turner, J. C. (1991). *Social influence.* Thomson Brooks/Cole Publishing Co.

Twitter (2021). https://transparency.twitter.com/en/reports/information-operations.html. https://transparency.twitter.com/en/reports/information-operations.html.

VOA (January 19, 2010). Russia, Libya sign $1.8 billion arms deal.

Walker, C. (2018). What Is 'Sharp Power'? *Journal of Democracy*, 29(3):9–23.

Walker, C. and Ludwig, J. (November 16, 2017). The meaning of sharp power: How authoritarian states project influence. *Foreign Affairs*.

Walsh, D. and Zway, S. A. (September 4, 2018). A Facebook war: Libyans battle on the streets and on screens. *New York Times*.

Walter, B. F. (2019). Propaganda and Radicalization in an Internet Age. *Available at SSRN 3477536*.

Watanabe, L. (2016). Libya's Future: Uncertain, Despite a Political Agreement. *Middle East Policy*, 23(4):114–122.

Wee, J. and Li, S. (2019). Politics and social media in the Middle East and North Africa. *Arab Barometer*.

Williams, G. I. (September 15, 2004). Journalism at a crossroads in Liberia. *NeimanReports*.

Yanagizawa-Drott, D. (2014). Propaganda and conflict: Evidence from the Rwandan genocide. *The Quarterly Journal of Economics*, 129(4):1947–1994.

Zaptia, S. (May 14, 2019b). HoR designates the Muslim Brotherhood a terrorist organization. *Libya Herald*.

Zaptia, S. (May 9, 2019a). Anti Haftar HoR faction holdes session in Tripoli. *Libya Herald*.

Zeitzoff, T. (2017). How social media is changing conflict. *Journal of Conflict Resolution*, 61(9):1970–1991.

Zhuravskaya, E., Petrova, M., and Enikolopov, R. (2020). Political effects of the internet and social media. *Annual Review of Economics*, 12.

# Appendix

## *Creating the dataset*

| Arabic and English search terms |
|---|
| أحمد المسماري طرابلس |
| عين زارة حفتر |
| العزيزية حفتر |
| طرابلس حفتر |
| عملية طوفان الكرامة طرابلس |
| بركان الغضب طرابلس |
| بركان الغضب حفتر |
| فايز السراج طرابلس |
| فائز السراج طرابلس |
| غزو طرابلس |
| هجوم طرابلس |
| حكومة الوفاق الوطني طرابلس |
| أحمد المسماري |
| عين زارة |
| فايز السراج |
| فتحي باشاغا |
| فائز السراج |
| غريان |
| مزدة |
| محمد قنونو |
| عملية طوفان الكرامة |
| بركان الغضب |
| قصر بن غشير |
| راس لانوف |
| صلاح بادي |
| Haftar Ain Zara |
| Tripoli Haftar LNA |
| Gharyan Haftar LNA |
| Gharian Haftar LNA |
| Tripoli Haftar Offensive |
| GNA Tripoli |
| LNA Tripoli |
| Fayez el Serraj Tripoli |
| Tripoli Ahmed al Mismari |
| Gate 27 Tripoli |
| Haftar Invasion |
| Haftar Offensive |
| Haftar Tripoli |
| Operation Flood of Dignity Tripoli |
| Operation Volcano of Anger Tripoli |
| Operation Volcano of Rage Tripoli |
| Salah Badi Tripoli |
| Wadi al Rabea Haftar |
| Wadi al Rabie Haftar |
| Yarmouk Military Camp Haftar |
| Zuwara Haftar |
| Ahmed al Mismari |
| Ain Zara |
| Ayn Zara |
| Azizia |
| Aziziyah |
| Bab al Azizia |
| Es Sider |
| Sidra |
| Fathi Bashagha |
| Fayez al Serraj |
| Fayez al-Sarraj |
| Gharian |
| Gharyan |
| Government of National Accord |
| Mizda |
| Mohammed Gnounou |
| Operation Flood of Dignity |
| Operation Volcano of Anger |
| Operation Volcano of Rage |
| Qasr ben Ghashir |
| Salah Badi |
| Yarmouk Military Camp |

**Figure 5.** Search terms used to create the dataset. We exclude search terms that did not return any results. While many of these terms on their own do not references Haftar's attack on Tripoli, in the month for which we pulled posts, these terms almost always referenced the attack. We manually read all of the posts and excluded those that were not about the attack.
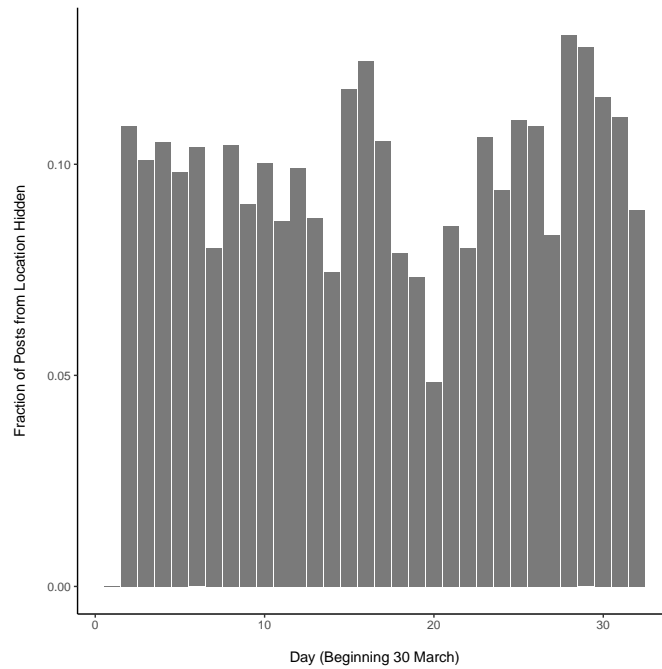
**Figure 6.** Fraction of posts originating from a Page with "Location hidden" over the course of the study period. This proportion appears to remain relatively stable over time. We note that the Page Transparency location information is a Page-level variable, not a post-level variable. The location(s) of administrator(s) will occasionally change, though this is not common. Given that Facebook is not completely transparent about how they assess administrator location, it is difficult to say why the location would change. It may change if an administrator moves permanently to a new country. We collected the administrator location data for all Pages at one point in time.

| GNA terms | LAAF terms |
|---|---|
| Coup | Restore |
| Invaders | Islamists |
| Legitimate | Liberation |
| Rogue | Extremists |
| Unbacked | Illegitimate |
| Renegade | Bravado |
| Warlord | Muslim Brotherhood |
| Siege | Brothers |
| Foreign fighters | Instability |
| Destabilization | Turkey |
| Rebels | Qatar |
| Man of war | Daesh |
| War criminal | al-Qaeda |
| Criminal | Militias Misratah |
| Mercenaries | Misrata militias |
| Strongman | Turkish invasion |
| Qarmatis | Syrian mercenaries |
| Militias Hfter | Libyan militias |
| Hjalh Corner | Field Marshal |
| Egyptian expansion | Turkish intervention |
| Rebel militias | Liberate the capital |
| Aggressor militias | Misurata Brigades |
| Counter-revolution | Islamic extremists |
| Aggression on the capital | Brotherhood leaders |
| Leviathan | The Government of the frigate |
| Wahabi heap | |
| City withstand | |
| Russian mercenaries | |
| Remnants Hfter | |
| Aggression against Tripoli | |
| Children plateau | |
| Attack Hfter | |
| Criminal Hfter | |
| Madkhali | |
| Halohta | |

**Table 6:** Dictionaries used to create slant measure. We selected these terms inductively in collaboration with a Libyan social media expert. We had the original English terms translated into Arabic by an Arabic speaker and then translated back to English with Google Translate, then added those terms to these dictionaries. While some of these terms are not themselves slanted, we include them because in the time period under study they were almost always associated with content that was slanted in a particular direction. For example, while Turkey is not itself a slanted term, it was almost always used to claim that the GNA was under foreign influence. Pro-GNA posts rarely referenced Turkey, wanting to understate the role of their foreign backers.

*Example Slant Measure*

We begin with an example post from our CrowdTangle dataset – this post on the Aljazeera Mubasher Channel is dated April 7, 2019.



We now walk through the processing steps we took to develop the GNA slant metric.

- Even though this post has an embedded video, there are no links to external websites so the *Description* field – which explains the content of external hyperlinks – is blank for this post in our CrowdTangle data.

- We use Google Translate to translate the text content of this Arabic post into English, producing "Urgent | A spokesman for the Libyan accord government forces: the launch of the process of a Volcano of Anger to cleanse Libya rebels." We do not translate posts that were originally written in English.

- We remove any URLs or special characters in the post, resulting in: "Urgent A spokesman for the Libyan accord government forces: the launch of the process of a Volcano of Anger to cleanse Libya rebels."

- We then break the text into unigrams and bigrams, stem all words, and remove entries with stop words. This produces the following:

  - Unigrams: *urgent, spokesman, libyan, govern, forc, launch, process, volcano, anger, cleans, libya, rebel.* Total number $= 12$

  - Bigrams: *govern forc, cleans libya, libya rebel.* Total number $= 3$ **Note:** we do not form bigrams from terms separated by stopwords; e.g. although "urgent" and "spokesman" are adjacent in the processed unigrams, they were not adjacent in the initial post. Without doing this, words that are separated by a string of stop words could unintentionally be matched to a two-word keyword.

- We now count the number of unigrams and bigrams that match processed terms from our two dictionaries:

  - Pro-GNA unigrams: *rebel* $= 1$

  - Pro-LAAF unigrams: none

  - Pro-GNA bigrams: none

  - Pro-LAAF bigrams: none

Our slant formula is:

$$Slant_i = \frac{GNA_i - LAAF_i}{Unigrams_i + Bigrams_i}$$

Inserting terms, our estimated slant for this post is $\frac{(1+0)-(0+0)}{12+3} = 0.067$, indicating a post that is estimated to be slightly slanted towards the GNA.

*Validating Slant Measure*

| Manual Code | Obs. | Mean | 95% C.I.: Lower | 95% C.I.: Upper |
|---|---|---|---|---|
| Strongly Pro-GNA | 140 | 0.0108 | 0.0066 | 0.0149 |
| Somewhat Pro-GNA | 88 | 0.0008 | -0.0014 | 0.0029 |
| Neutral | 89 | 0.0014 | -0.0006 | 0.0033 |
| Somewhat Anti-GNA | 53 | -0.0000 | -0.0017 | 0.0016 |
| Strongly Anti-GNA | 93 | -0.0084 | -0.0131 | -0.0037 |

**Table 7:** We show the number of observations, mean value of *raw slant*, and 95% confidence interval for each level of the manual coding conducted by the independent coder on a random sample of 500 posts from the data set. Because our slant metric is a continuous measure of relative slant (and not a binary classification of pro-LAAF and pro-GNA posts), we report these summary statistics in support of the conclusion that this measure captures relative slant. As shown above, the slant measures for posts that are hand-coded as "Strongly Pro-GNA" are more pro-GNA (larger, in a positive direction) than the posts that are hand-coded as "Somewhat Pro-GNA; likewise, the posts that are hand-coded as "Strongly Pro-LAAF" are more pro-LAAF (larger, in a negative direction) than posts that are "Somewhat Pro-LAAF." We take this as evidence that the slant metric is capturing the relative degree of pro-GNA or pro-LAAF language in a post. The one caveat is that *neutral* posts – as assessed by the independent coder – are determined by the *slant* measure to be slightly more pro-GNA than coder-assessed "Somewhat Pro-GNA" posts. We investigated this, and observed that the independent coder did not pick up on some coded language, such as referring to Haftar as a "rebel" and anti-GNA groups as "militias" – both are pro-GNA terms. Picking up on this language is a feature of our automated strategy, which involved careful selection of slanted terms which may not be obviously slanted to someone who is not familiar with Libyan social media.
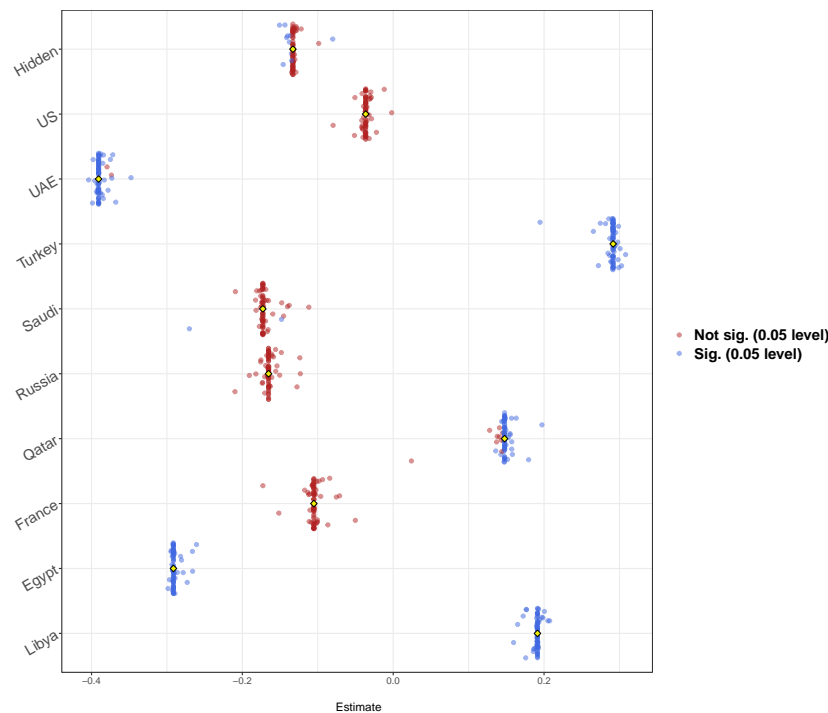
**Figure 7.** This figure shows the sensitivity of the regression estimates of *standardized slant* on post location (column 1 from Table 3) to the specific terms included in our GNA and LAAF terms (found in Table 6). The yellow diamonds indicate the point estimate for each country location, as reported in Table 3 of the main text. The points show the regression estimates – following the same specification as in column 1 of Table 3 – from each of the dictionaries that can be formed from dropping one term found in Table 6. Estimates that are statistically significant at the 0.05 level are colored in red; estimates that are not statistically significant at the 0.05 level are colored in blue. The results from Table 3 are robust to dropping dictionary terms: the estimates for Libya, Egypt, Qatar, Turkey, and the UAE are consistently in the same direction as in 3 and statistically significant at the 0.05 level, regardless of the term dropped from the dictionary. This evidence supports the conclusion that the method of assessing slant used in this paper is not sensitive to specific words or phrases.

*Geographic distribution of Facebook posts and slant*
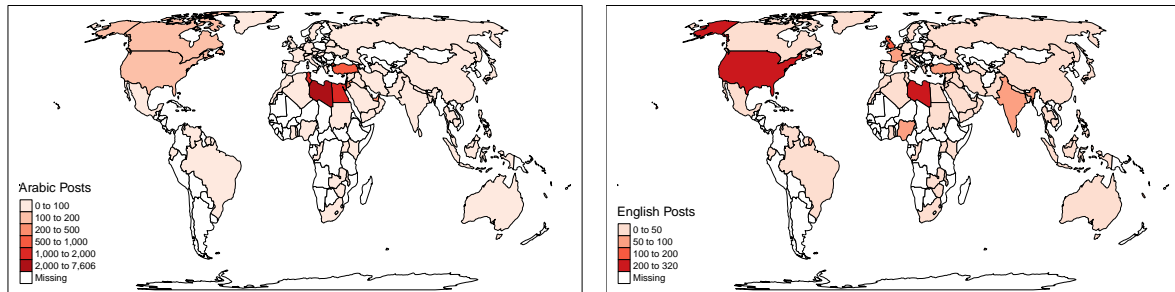


**Figure 8.** Geographic distribution of Facebook posts by language (Arabic, English). Note that color labels correspond to different magnitudes in each panel.
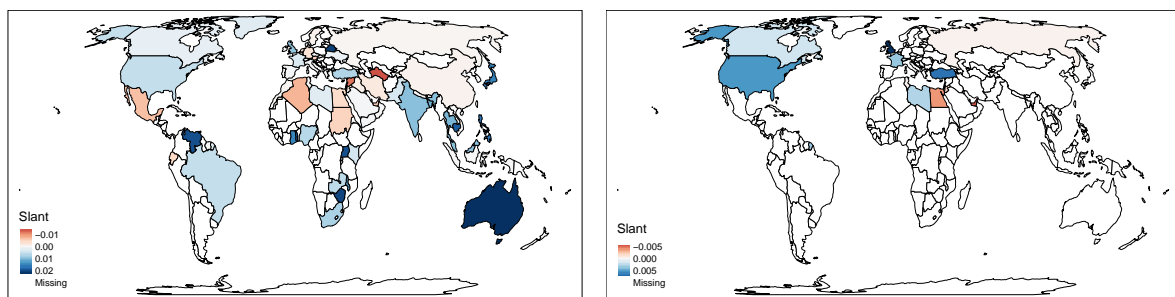


**Figure 9.** Left: Average slant level for all countries in data set. Higher (more blue) slant indicates more pro-GNA bias. Right: Average slant level, only for countries with at least 100 posts.

*Facebook posts by day*
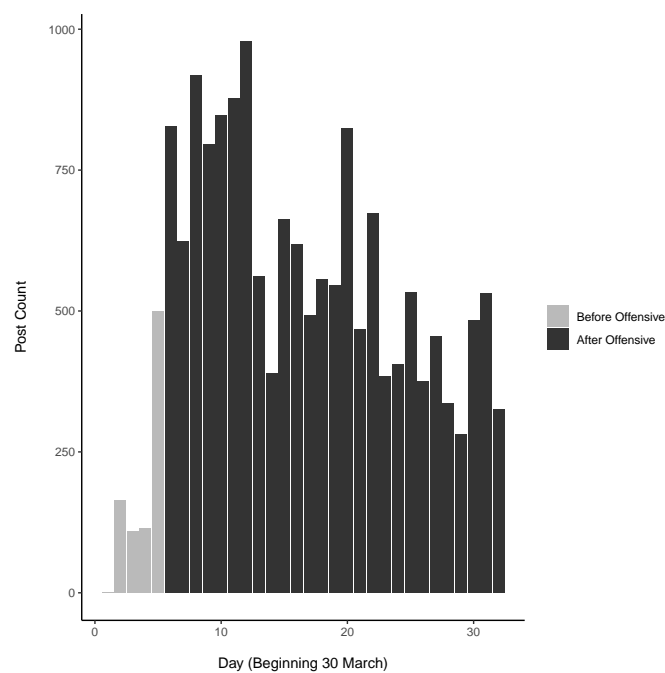


**Figure 10.** The number of Facebook posts per day during the study period (March 30–April 30). The number of posts increased during the period immediately following the Tripoli offensive (April 4) and remained relatively high–compared to the number of posts before the offensive–throughout the rest of April. Note that the time of a Facebook post is recorded in the data using Pacific Time, hence some posts may actually be on a different day in their local time zone; this is one possible reason for the rise in the number of posts observed on the day prior to the beginning of the offensive. Another possibility is that some Page administrators were aware that the offensive was about to happen.

*Day and slant*
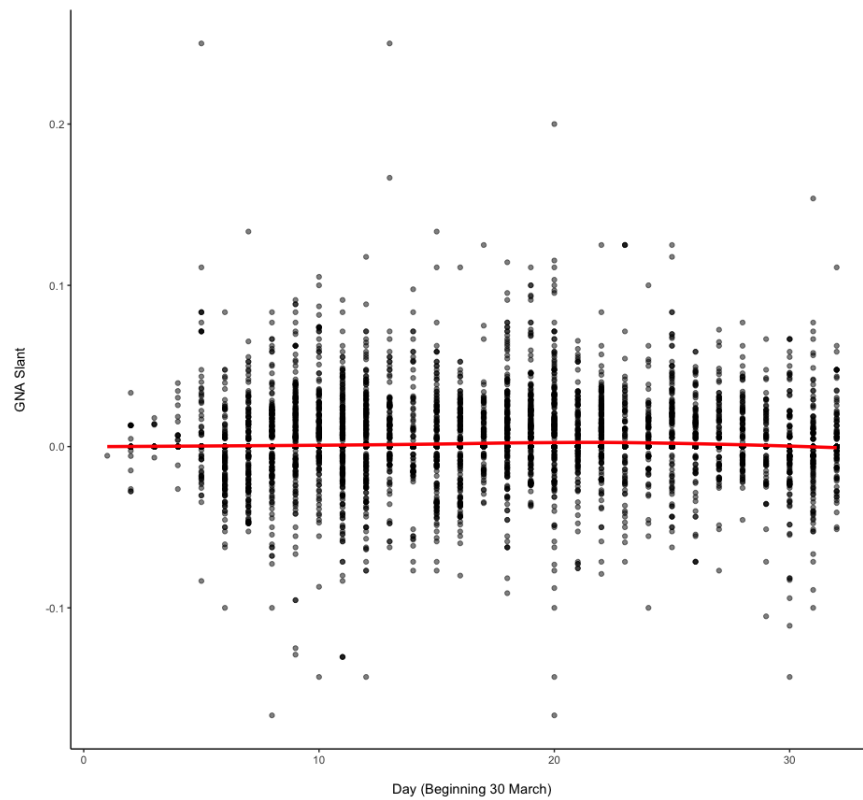


**Figure 11.** Raw distribution of GNA slant by the day of posting. The red line super-imposes a LOESS curve. There does not appear to be a meaningful relationship between GNA slant and the day of the month, suggesting that there is no trend in partisan language over the study period.

### *Language and slant*

|  | Model 1 | Model 2 |
|---|---|---|
| English | 0.41*** | 0.40*** |
|  | (0.06) | (0.08) |
| Location Fixed Effects? |  | Yes |
| R$^2$ | 0.02 | 0.05 |
| N. Observations | 16662 | 16662 |
| N. Unique Pages | 1487 | 1487 |

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

**Table 8:** Coefficients from regressions of *GNA slant* on an *English language* indicator. Model 2 includes Facebook Page location fixed effects. Standard errors clustered at the Facebook Page level are included in parentheses.

**Distribution of raw slant by country**
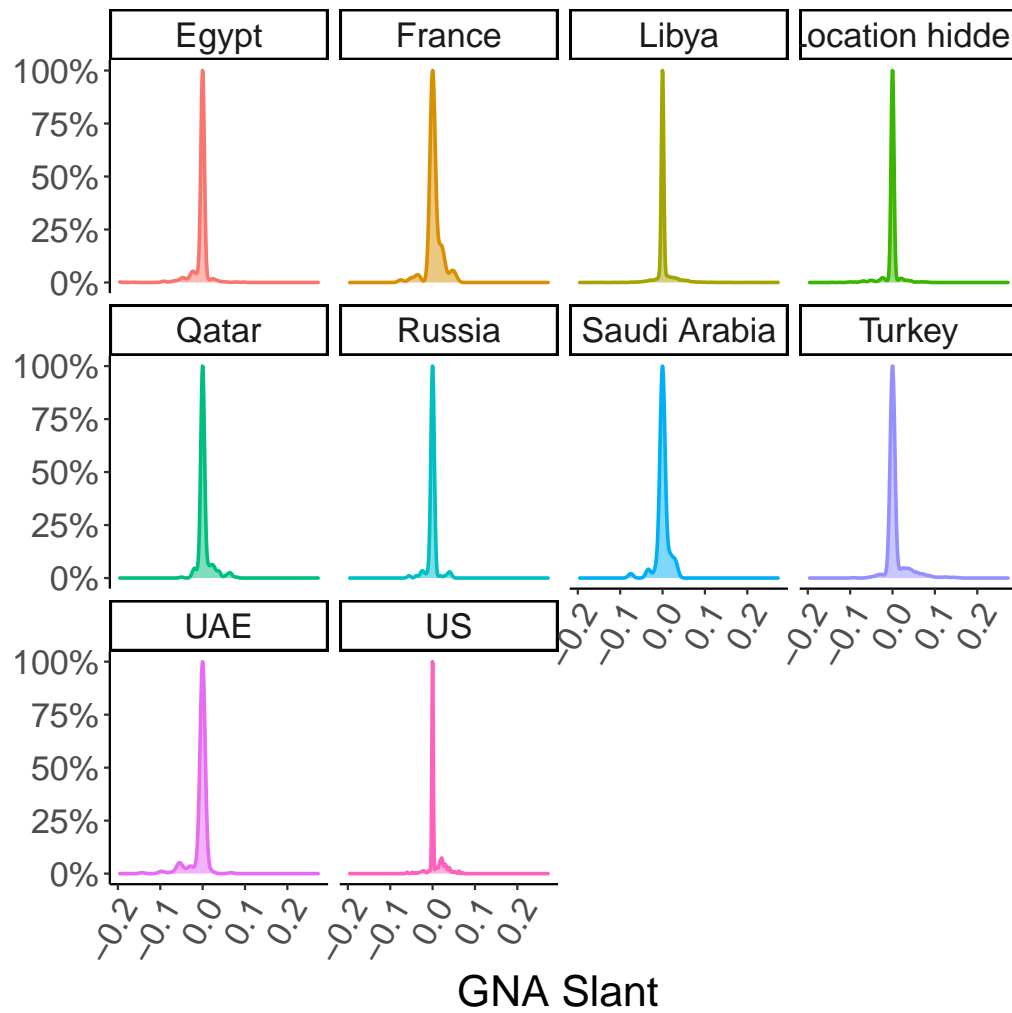


**Figure 12.** Density plots for the non-normalized, raw GNA slant by country of origin. Across all countries, posts are clustered around zero slant.

*Regressions that include all country indicators*

|  | Standardized Slant | LAAF Count | GNA Count |
|---|---|---|---|
| Libya | 0.17*** | −0.05 | 0.13*** |
|  | (0.04) | (0.03) | (0.03) |
| Egypt | −0.18*** | 0.15* | −0.04 |
|  | (0.06) | (0.08) | (0.03) |
| France | −0.04 | −0.01 | −0.01 |
|  | (0.11) | (0.07) | (0.08) |
| Qatar | 0.22** | −0.09** | 0.12 |
|  | (0.06) | (0.03) | (0.10) |
| Russia | −0.09 | −0.06 | −0.14 |
|  | (0.15) | (0.04) | (0.11) |
| Saudi Arabia | −0.11 | 0.08 | −0.08 |
|  | (0.08) | (0.13) | (0.05) |
| Turkey | 0.35*** | −0.12*** | 0.09 |
|  | (0.11) | (0.03) | (0.06) |
| UAE | −0.31* | 0.08 | −0.12** |
|  | (0.15) | (0.08) | (0.03) |
| USA | 0.02 | 0.01 | 0.04 |
|  | (0.06) | (0.04) | (0.05) |
| Location Hidden | −0.04 | −0.05 | −0.04 |
|  | (0.06) | (0.03) | (0.03) |
| Language Fixed Effects? | Yes | Yes | Yes |
| $R^2$ | 0.04 | 0.01 | 0.02 |
| N. Observations | 16662 | 16662 | 16662 |
| N. Unique Pages | 1487 | 1487 | 1487 |

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

**Table 9:** Coefficients from three regressions of the outcomes – *Standardized GNA Slant*, *LAAF Count*, and *GNA Count* – on a set of relevant country indicators are reported above. These regressions control for language $\in$ (Arabic, English fixed effects. Standard errors clustered at the Facebook Page level are included in parentheses. Each regression is estimated on a set of 16,662 observations across 1,487 unique Facebook Pages.

### *Engagement*

Table 10 shows post engagement by country. We note that these statistics should be interpreted with caution. In countries with repressive governments and militias, such as Libya, social media users may be reluctant to publicly engage with content.

| Country | Number Posts | Mean Engagement | Mean Likes | Mean Shares | Mean Comments |
|---|---|---|---|---|---|
| All posts | 16662 | 570.555 | 476.301 | 13.802 | 50.909 |
| Libya | 7926 | 658.687 | 571.187 | 9.837 | 49.342 |
| Egypt | 1685 | 311.282 | 258.123 | 15.833 | 22.037 |
| Location hidden | 1596 | 795.169 | 639.765 | 18.031 | 91.878 |
| Turkey | 790 | 637.99 | 530.549 | 17.611 | 55.41 |
| USA | 378 | 232.741 | 124.693 | 37.511 | 23.608 |
| Qatar | 298 | 631.718 | 450.299 | 64.121 | 64.99 |
| UAE | 266 | 559.256 | 472.481 | 11.568 | 47.921 |
| France | 113 | 151.575 | 105.54 | 8.956 | 23.77 |
| Russia | 105 | 555.371 | 350.581 | 38.771 | 117.61 |
| Saudi Arabia | 64 | 124.75 | 94.484 | 5.812 | 17.312 |

**Table 10:** Descriptive statistics on engagement figures are included for all posts and select countries (the same set found in Table 2. *Engagement* is calculated from the sum of *Likes*, *Comments*, *Shares*, and the "Love", "Wow", "Haha", "Sad", "Angry", and "Thankful" reactions.

| | Engagement | | Likes | | Shares | | Comments | |
|---|---|---|---|---|---|---|---|---|
| \|Standardized Slant\| | 2.62 | −2.78 | −8.03 | −13.94 | 3.88* | 4.80* | 4.15* | 2.90 |
| | (15.67) | (18.25) | (12.39) | (14.39) | (2.03) | (2.60) | (2.47) | (2.03) |
| log(Page Likes) | | 126.89*** | | 98.61*** | | 5.19*** | | 13.51*** |
| | | (20.36) | | (16.57) | | (1.11) | | (2.33) |
| Language FE? | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Country FE? | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $R^2$ | 0.07 | 0.10 | 0.08 | 0.11 | 0.01 | 0.01 | 0.05 | 0.07 |
| N. Observations | 16662 | 14460 | 16662 | 14460 | 16662 | 14460 | 16662 | 14460 |
| N. Unique Pages | 1487 | 1285 | 1487 | 1285 | 1487 | 1285 | 1487 | 1285 |

$^{***}p < 0.01; ^{**}p < 0.05; ^{*}p < 0.1$

**Table 11:** Coefficients from regressions of social media interaction on the standardized slant of the post. |Standardized Slant| is the absolute value of the standardized GNA slant measure. The *log* of page likes at the time of posting is included as a control variable. The number of page likes at the time of posting is missing for 2,202 posts. *Engagement* is calculated from the sum of *Likes*, *Comments*, *Shares*, and the "Love", "Wow", "Haha", "Sad", "Angry", and "Thankful" reactions.

### *Misclassified Page*

A manual inspection of the 20 most slanted Pages revealed that a Page classified as pro-LAAF was not, in fact, pro-LAAF. While we feel confident in our dictionary method based on the validation exercise, we investigated this Page's posts to assess whether we may have missed a systematic problem with the dictionary. One post was categorized as pro-LAAF because it used the word liberation, a word that Haftar's camp uses much more commonly. Another post referred to Haftar's "Libyan Army". That is generally seen as a pro-LAAF phrase, as the LAAF sees themselves as the legitimate army. But in this rare case a Page that is not pro-LAAF appears to be using it. These instances seem to be idiosyncratic, and were not systematic enough to merit adjusting the dictionaries.

### *How do social media users respond to slanted content?*

The posts in this dataset received a considerable amount of engagement, with an average of approximately 570 interactions per post (see Table 10 in the Appendix). Though there is no indication in our data that social media consumers engaged more – as measured by total interactions (reactions, comments, and shares combined) – with slanted posts, there is some evidence that slanted material in both the pro-GNA and pro-LAAF directions increased the number of times the post was shared and the number of comments on the post (see Table 11 in the appendix). To better understand how users interact with slanted material, we conducted a deep dive on some of the most slanted posts.

Specifically, by looking at how social media users respond to a subset of slanted posts, we can make some assessments as to whether citizens ever respond skeptically to such content. To do this, we looked at the first ten comments on the ten most LAAF-slanted posts and the first ten comments on the ten most GNA-slanted posts. We coded each comment as "affirmative" (the commenter agreed with the post), "critical" (the commenter disagreed with the post), "ambiguous" (the slant was not obvious and difficult to assess) or "neither" (for example a user tagging a friend in a comment). We note that this relatively small-N analysis should be interpreted as only suggestive. Overall, we find that 47% of the comments on these slanted posts are affirmative and 13% are critical. Critical comments were more frequent in the pro-LAAF posts: 17% of comments on these posts were critical, compared to 8% on the pro-GNA

posts (see Table 12).[25]

|  | Pro-LAAF Posts | Pro-GNA Posts |
|---|---|---|
| Affirmative | 58% | 37% |
| Critical | 17% | 8% |
| Ambiguous | 4% | 37% |
| Neither | 20% | 18% |

**Table 12:** Commenter Reactions on Slanted Posts

For example, one pro-LAAF post said, "Libya Dawn. Volcano of Anger. They are a group of terrorist militias driven and controlled by Islamic groups such as Brotherhood, Fighter, Jihadists and others" (translated). Libya Dawn is a group of pro-Islamist militias and Volcano of Anger is the name of the GNA counteroffensive to regain territory captured by Haftar's forces. One user commented: "Praise be to God for your safety, you beautiful ones" (translated). We code this as an affirmative comment as it is supporting LAAF forces fighting GNA aligned forces.

As another example, one anti-GNA post said, "Salah Badi, is the leader of the so-called terrorist 'al-Samoud Brigades' supported by Turkey and Qatar." Salah Badi is an internationally sanctioned commander of the al-Samoud Brigade, a militia affiliated with the GNA.[26] Haftar's supporters, including the governments of Saudi Arabia and the UAE, use this alliance to critique the GNA. One user commented: "The leader of terrorism is Haftar, who is backed by the Saudi and Emirati Arab Zionists" (translated). We code these types of comments as critical.

Many comments were neither affirmative nor critical. For example, one post said: "God is the greatest, fusion/coalescence of the revolutionaries of Tajura and the rebels of the Ain Zara axis." Tajoura and Ain Zara are suburbs southeast of Tripoli experiencing intense violence

---

[25]We note that it is possible that some portion of these comments are from individuals working for a covert influence operations. This tactic has been observed in the past.

[26]washingtonpost.com/world/middle_east/he-once-attacked-tripoli-now-a-libyan-militia-leader-defends-it-from-another-invader/2019/07/19/52c8a0b8-a258-11e9-a767-d7ab84aef3e9_story.html

amid the advancing LAAF and defending GNA. This post praises Haftar's forces. One user commented "Bad thieves in Libya [thinking face emoji]" (translated). We code this and similar comments as neither, given ambiguity as to whether the commenter is calling the GNA and its associations militias thieves, or Haftar's forces thieves.

That 13% of comments on the most slanted posts in our dataset were critical suggests that social media users do not always passively absorb biased content. Though we caution that the sample size for this analysis is not large. Future research could consider the conditions under which users are more or less likely to respond critically to content.

### *Page audience*

Analysis of these Pages is incomplete without an understanding of the location of users active on these Pages. While Page administrator location is useful, the Page Transparency feature does not provide data on the consumers of the content. To identify users' locations we randomly sampled 20 posts from the dataset, and randomly sampled five comments from these posts. This gave us a dataset of 83 comments (some posts had fewer than five comments) and unique commenters.[27] We then set out to identify the location of these commenters. To do this we looked at the "About" sections of their accounts and the location tagged in posts. For several accounts, we inferred location from the university the respondent attended.

We find that of the 83 commenters, 56 live in Libya, two in Tunisia, and one in: Egypt, Denmark, Czech Republic, and Saudi Arabia. For 21 commenters we were not able to ascertain their location. This suggests that despite the prevalence of Page administrators from outside Libya, it appears that a majority of users interacting with these Pages are in Libya. We note, however, that the sample size for this analysis is small, and our ability to confidently assess user location is limited. Additionally, users who appear to be in Libya may not be authentic accounts.

---

[27]Some posts had fewer than five comments. This is why we have fewer than 100 comments.