# Opinion – Twitter as an Orwellian Global Editor

https://www.e-ir.info/2023/06/19/opinion-twitter-as-an-orwellian-global-editor/

NOOR SUWWAN, JUN 19 2023

Twitter has been singled out by scholars studying the intersection of politics and cyberspace for its unmatched political momentum, amid other communication platforms. Consider for instance that as of 2020, 98% of the 193 UN member states had an active Twitter presence, represented by 163 heads of states and 136 Foreign Ministries (Twiplomacy 2020). So self-evident is the clustering of a digital international community on Twitter that identifying the absent few is a relatively easy task. For example, missing in 2020 were only four states, namely, Laos, North Korea, Sao Tome and Principe and Turkmenistan (Twiplomacy 2020).

According to Marc Lynch (2011), the novelty of digital communication technologies lies in their construction of a new route to information sharing (Lynch 2011). By that, bypassing information producers, previously monopolized by large media corporations. To him, a voice in the information ecosphere is so critical that digital platforms are actually tools of political combat (Lynch 2011). Hence, it was access to these tools that enabled the carving out of a new public sphere, in an Arab context (Lynch 2007).

Within the new digital environment, Twitter effectively occupies a unique sociopolitical position (Collins, DeWitt and LeFebvre 2019; Golbeck, Grimes and Rogers 2010; Lewandowsky, Jetter and Ecker 2020; Manor and Segev 2020; Pavón-Guinea 2018). For instance, Lieu et. al (2010) found that information sharing and seeking are two of the most fundamental motivators for Twitter's user base (Liu, Cheung and Lee 2010). Similarly, the American Press Institute found that Twitter users significantly consume more news than users of other social media sites (Rosenstiel, et al. 2015). Furthermore, according to the Geneva-based public relations firm, Burson-Marsteller, governments around the world also use Twitter mainly for information-sharing (Twiplomacy 2020), enhancing Twitter's perceived efficacy in information circulation.

It is this perceived efficacy in information circulation that triggers significant spikes in new accounts during events of geopolitical relevance by a global public sphere eager to weigh in. For instance, the Indiana University observatory on social media recorded a peak in the number of new accounts on Twitter around the day of the Russian invasion of Ukraine (Indiana University's observatory on social media 2022). Similarly, the 20222 French presidential elections also saw a peak in new account proliferation (Pierri, Luca and Ferrara 2022), quantifying a growing universal culture of claiming a stake in the information ecosphere.

It this information ecosphere that Twitter, too, decided to play an active role in, as of 2020. It started with an emergency verification campaign, to identify health experts citing concerns of misinformation around the Coronavirus pandemic. That soon expanded to a general verification of political actors. However, this digital diplomatic recognition did not unravel evenly amongst the community. For instance, Botswanan President, Mokgweetsi Masisi who tweeted celebratorily after having received his Instagram verification and said *"…working hard with my team to ensure that Facebook and Twitter also verify the other accounts"* has yet to receive his blue badge; the visual marker of a digital recognition (Twiplomacy 2020). So is the Czech Foreign Minister, Tomáš Petříček (Twiplomacy 2020). Meanwhile, the Czech Foreign Ministry had to publicly plea for verification from Twitter ironically in a tweet, a request that was only to be granted several months later (Twiplomacy 2020). For better or worse, Twitter had effectively become a new recognizing body, albeit an online one. With that, actively interfering in the actualization of a modern state's personhood.

## Opinion – Twitter as an Orwellian Global Editor

Written by Noor Suwwan

After its campaign against misinformation, Twitter moved on to policing disinformation. For instance, in January 2020, Twitter suspended several accounts linked to the Venezuelan government, including Nicolas Maduro's own presidential account (Twiplomacy 2020). It also suspended Venezuelan Vice President, Delcy Rodríguez's personal account while her institutional account was restricted, citing 'unusual activity' (Twiplomacy 2020). These measures fall under the category of *hard moderation* of Twitter's content moderation policy. Hard moderation involves affirmative action against content deemed harmful that spans message deletion, temporary account suspension or user de-platforming (Zannettou 2021).

After years of claiming inability to reliably remove terrorist content citing unfeasibility of doing so without engaging in political value judgements (Díaz and Hecht-Felella 2021), Twitter's new hard moderation policies present a stark departure from its earlier apolitical commitment. For example, QAnon has had a long history of active Twitter use for information circulation, with minimal intervention on Twitter's behalf, despite violations of Twitter's well-articulated hateful conduct policy (Díaz and Hecht-Felella 2021). However, after the January 6 U.S Capitol, Twitter passed its 'coordinated harmful activity' policy, under which it, for the first time, permanently suspended thousands of QAnon associated accounts (Díaz and Hecht-Felella 2021).

Similarly, succumbing to pressure mainly from American and European governments to remove so called terrorist content, Twitter expanded its 'violent organizations policy', under which the demanded content could be removed (Díaz and Hecht-Felella 2021). While this is inherently predicated on identifying groups categorized as terrorist organizations, Twitter has yet to publish which groups it labels as such. This means that users do not necessarily know when they are engaging in rhetoric that can be translated as support of a 'violent organization'. It is this insistence in engaging in imprecise language that enabled Twitter's moderation policies to disproportionately target Muslim and Arabic-speaking communities (Díaz and Hecht-Felella 2021).

Twitter's linguistic imprecision also fuels public perceptions of a political bias to content moderation, amounting to censorship claims in some cases (Zannettou 2021). Zannettou (2021) investigated accusations of a Twitter political bias and indeed found evidence of disproportionate moderation of Republican content, as opposed to Democrat content, by a factor of 7 (Zannettou 2021). Similarly, Twitter complied with 82% of requests submitted by the Israeli Attorney General's office to remove pro-Palestinian content during the May 2021 Israeli-Palestinian conflict escalation (Díaz and Hecht-Felella 2021), exposing not only an unintentional bias but an active affinity to some political ideologies.

Further problematizing Twitter moderation is scholarly research that concludes the presence of major inconsistencies in action against policy violations (Zannettou 2021) (Paudel, et al. 2023). For example, similar content worded differently weren't always moderated equally (Zannettou 2021). Similarly, content in English was more likely to be moderated than other languages (Zannettou 2021). Furthermore, the majority of moderated content in 2020 were in relation to the American presidential elections (Zannettou 2021); resulting in the suspension of 2 million accounts (Pierri, Luca and Ferrara 2022), suggesting a US-centric content moderation strategy.

Twitter's transit into a more decisive content moderator enables it to carve out a role for itself in International Relations; that of a global editor. With that, disarming the global public sphere of an efficient tool for political combat and threatening Twitter's status as an alternative route to information sharing. Instead, Twitter's current editorial role suggests the reproduction of geopolitical power dynamics, with its increasingly visible bilateral cooperation with states and their (political) requests. This allows Twitter's fusion with old media corporations, as present and past information factories. Similarly, Twitter's hard moderation policies amount to one familiar Orwellian nightmare; that of *vaporization*. Vaporization is firing content out of existence. In its most extreme form, vaporization can amount to *unpersoning*; revoking the existence of a person. Twitter's accelerating routinization of content deletion models an Orwellian vaporization, while its account suspension and de-platforming digital unpersoning. So much so that Twitter is now ideally positioned to vaporize and unperson entire states, if prompted.

### References

Collins, S.D., J.R. DeWitt, and R.K. LeFebvre. 2019. "Hashtag Diplomacy: Twitter as a tool for engaging in public

diplomacy and promoting US foreign policy." *Place Brand Public Diplomacy.*

Díaz, Ángel, and Laura Hecht-Felella. 2021. *Double standards in social media content moderation .* Brennan Center for Justice at New York University School of Law.

Indiana University's observatory on social media. 2022.*Analysis of Twitter accounts created around the invasion of Ukraine.* Indiana University .

Liu, Ivy L.B., Christy MK Cheung, and Matthew K. O. Lee. 2010. "Understanding Twitter Usage: what drive people to continue using Twitter." *Pacific Asia Conference on Information Systems.*

Lynch, Marc. 2011. "After Egypt: the limits and promise of Online challenges to the Authoritarian Arab state." *Perspective on Politics.*

—. 2007. *Voices of the New Arab Public: Iraq, Al Jazeera and Middle East Politics today .*

Paudel, Pujan, Jeremy Blackburn, Emiliano De Cristofaro, Savvas Zannettou, and Gianluca Stringhini. 2023. "Lamberrta: learning to rank for Twitter soft moderation." *44th IEEE Symposium on Security & Privacy.*

Pierri, Francesco, Luca Luceri, and Emilio Ferrara. 2022. "How does Twitter moderation work? Dynamics of account creation and suspension during major geopolitical events." *Association for the advancement of Articial Intelligence.*

Rosenstiel, Tom, Jeff Sonderman, Kevin Loker, Maria Ivancin, and Nina Kjarval. 2015.*Twitter and the News: How people use the social network to learn about the world.* The American Press Institute.

Siegel, Alexandra A., Jonathan Nagler, Richard Bonneua, and Joshua A. Tucker. 2021. "Tweeting beyond Tahrir: ideoligical diversity and political intolerance in Egyptian Twitter networks." *World Politics.*

Twiplomacy . 2020. *Twiplomacy Study 2020.* Twiplomacy .

Zannettou, Savvas. 2021. ""I won the Election!": An Empirical Analysis of Soft Moderation Interventions on Twitter." *Proceedings of the Fifteenth International AAAI Conference onWeb and Social Media.*

---

**About the author:**

**Noor Suwwan** is an Adjunct Faculty of International Relations at the University of Wollongong in Dubai. Her main research interests are what she calls 'Cyber-International Relations' and the Political Science of Algorithms.