ASCE
ARTA STUDY CENTRE FOR EUROPE UNIVERSITY OF KARACHI

https://asce-uok.edu.pk/journal
DOI: https://doi.org/10.56384/jes.v41i1.365

The Role of Anonymity in the Weaponization of Social Media through Twitter: Some Global and European Instances

Sarah Syed Kazmi*

Abstract

Digital social media platforms have been instrumental in influencing the global ecosystem of popular interaction. The article will delve into the role of anonymity as a stimulus abetting aggressive behaviour particularly on Twitter. Trolling and other detrimental societal patterns result from a striking disconnect between actors posing anonymity and those they address. The anonymous indulgence in pejorative comments, derogatory dialogues, and multifarious patterns of hate speech precipitates into a weapon in the hands of terrorists and mala fide actors functioning on Twitter. Whether AI sentiment analysis can act as a viable check to regulate inappropriate behaviour will be addressed in the light of theories on human freedom as this might furnish third-party filters, oversight surveillance by moderators averse to proponents of online freedom. Also, increased levels of monitoring, enforcement of protective regulations may constrict the content and may appear as a breach of free speech. The role of European Union in addressing challenges posed by the proliferation of social media platforms will be dilated upon to check the efficacy of these regulating practices.

Keywords: Anonymity, Free speech, Hate speech, Social media, EU policy

Introduction

The major objective behind social media is to bring people together in the global social milieu. Social media platforms collect, accumulate, store, share, deliberate, and deliver user generated and general media content which influences popular perception directly or indirectly. Social media platforms offer divergent narratives to the target audience causing an onslaught of information. In case a society is witness to the 'oppressor' and 'oppressed'

* Dr. Sarah Syed Kazmi is a Director, Quality Enhancement Cell (QEC), and Head of the English Department at Fatimiyah Higher Education System, Karachi. Email: sarahkazmi@fhes.fen.edu.pk.

dichotomy, social media becomes an online ground of warring narratives, further deepening the fault lines of 'us' versus 'them'. X, formerly twitter, is a famous social media platform which allows digital communication and dissemination of speech. Patterns of polarization where retweets in the affirmative endorse a certain viewpoint are seen as a natural outcome of being attracted to similarity and repulsed by stark differences.[1] However, it becomes problematic when social groups operating on opposite poles cease to engage in a non-violent manner and the communication tilts in the direction of hate speech.

Twitter was launched in 2006. It allowed posting short messages comprising 140 characters; a form of microblogging. By default, these short messages or tweets are publicly visible which do not mandate having a Twitter account to read them. X account, (formerly Twitter), can be formed by creating a profile which gives a general overview about the user. This is inclusive of the username and name/last name (could be user's actual name or a pretended identity). Sometimes a URL which links to another social media platform is also given. After account activation, users can 'tweet', i.e. post a message which does not exceed 280 characters as of now, including 'hashtags'. As a user follows a 'friend', he/she will be able to view his/her tweets. This is not reciprocal, i.e. the 'friend' might not follow the one following him/her and will not be able to receive his/her tweets. Users, who follow a particular account, are known as 'followers' and they remain posted about recent tweets of their friends. Due to absence of stringent measures many accounts operate under fake identities promoting anonymity.

Anonymity is usually masqueraded in the guise of a pseudonym, despite the fact that creating an account over X does require an email address and, at times, one's mobile number for authentication. These pseudonyms are usually distinct from the real identities of users as intended, while some are adopted for the sake of fun. The relationship between content sensitivity and user anonymity is directly proportional and also one of the factors behind the use of anonymous identities. Anonymity results in a pronounced deindividuation implying greater liberty in terms of online interactions. While the individual identity is preposterous, it ironically results in strong group affiliations. As the locus shifts from self, newly concocted identities, generic titles, and group conformity become more salient. Thus, anonymity

[1] Peter Coleman. *The Way Out: How to Overcome Toxic Polarization* (New York: Columbia University Press, 2021), 221.

generates a detachment for users to disassociate themselves from their salient, online behavioural pattern and evade responsibility in case of an overt anti-social expression. It can be argued that anonymity abets uncivil discursive practices rendering targets of hate speech vulnerable and resulting in a directly proportional relationship between anonymity and sensitive content posted with impunity.

Sensitive content could be further bifurcated as one which correlates with anonymity and the other wherein users proclaim their identities despite content sensitivity. For example, accounts pertaining to the use of drugs and physical orientation are found to be largely anonymous as a measure to evade judgment and censure while those promoting racism, apartheid, and belligerence are usually open and not clandestine. Yet, it is interesting to note that even regular users choose to remain anonymous, reaffirming that users do not always create anonymous profiles to express opinions about sensitive issues. People with distinct and divergent opinions either seemingly align with the majority or may choose to remain silent for the fear of isolation. Aligning with the majority is a convenient means of legitimizing one's opinion yet those with deep-seated, divergent, and differing opinions choose to remain silent or voice their concerns through feigned or anonymous identities.

It is a matter of perceived affordances with respect to X where fear of isolation makes it convenient for users unwilling to take ownership of posted content and bask in anonymity. Network affiliation, social presence, and anonymity offer interesting insights with respect to different social media platforms. For example, on Facebook due to its operational pattern and the liberty of sharing personal information on the wall, lesser degree of anonymity results in conscious efforts aimed at self-censorship and polite discourse.[2] On X, owing to greater chances of clandestine identities, open discussions on controversial topics take place without fear of chastisement from opponents.

Typically, online platforms ask for an online identity by mandating creation of account and adoption of a virtual identity. However, there is a disparate approach on part of the online platforms; some advocating privacy under the pretext that real identities mar users' freedom by compelling one to share

[2] Sauvik Das and Adam Kramer. "Self-Censorship and Facebook", at https://ojs.aaai.org /index.php/ICWSM/article/view/14412.

personal information with a large number of online users and making them increasingly vulnerable to trollery. Many social media platforms like Facebook and Google related accounts call for a real-name policy at the time of creation of accounts. One of the most cited reasons is that it increases the onus of responsibility on users in terms of the content shared, thereby ensuring accountability. Since the policy is not strictly enforced, therefore, many users conveniently resort to pretended identities.

A cross-sectional analysis of the user identities shows that users can be categorized as those who are 'identifiable', 'partially identifiable', and 'anonymous'. The anonymous users are proactive in showcasing their activities, voicing opinions, and hence the greater number of their tweets. They lurk less and actively follow accounts. Sometimes anonymity is practised in order to conveniently follow controversial accounts as users' perceived freedom increases with covert identities. Consequently, the online persona may be entirely different from the user's offline image and perception of selfhood. Anonymity, therefore becomes a tool of free speech be it positive or hate speech.

It can be derived from web-based studies based on two filtering questions (the first question inquires whether users use X at least once a week; the second question is thematic, eliciting whether they lately expressed their opinion on socio-political issues i.e. once in a week), exhibit interesting relationship between anonymity and political engagement. The respondents are further filtered as active if they have more than 100 followers. The engagement levels demonstrate that users' higher anonymity leads to higher engagement level especially with respect to political comments. This can be further testified as users with high anonymity level and higher frequency of political comments show higher degree of engagement level. Users with lesser degree of anonymity corresponds to a reticence in expressing opinions regarding political comments. This relationship indicates that higher anonymity correlates with higher engagement levels. Engagement also increases with political themes, where users who comment frequently on political topics tend to have higher overall engagement.

Global Impact on Geopolitics

With the fast-changing contours of the virtual world, the use of laptops, tablets, and gaming consoles is also on the rise.[3] Access to smartphones which leads to multitudinous social media platforms has become a 'rite of passage' for adolescents. This democratization of technology brings certain challenges such as trolling, deep fakes, use of pejoratives, derogatory posts, and an onslaught of disinformation propelled by equally anonymous and fake accounts primarily due to lack of stringent regulation. The ever-evolving virtual world, with hundreds of new users a second, is in a state of constant flux. For example, as per X's CEO's declaration, Twitter boasts of more than 500 million active users a month.[4] This brings a huge amount of information to equally large audiences instantaneously and, at times, anonymously which may pose serious security threats as non-state actors and terrorists employ a hybrid strategy of cyberattacks and a (dis)information campaign. This is further intensified by the sheer volume of users' traffic surfing social media platforms coming directly into contact with concocted information. Users, mass mobilizers, and political actors take into account stimuli driving media fury, steering it further to achieve radical agendas or simply sustain international attention. In this way, the underrepresented groups remain on the verge of online abuse as in real life.

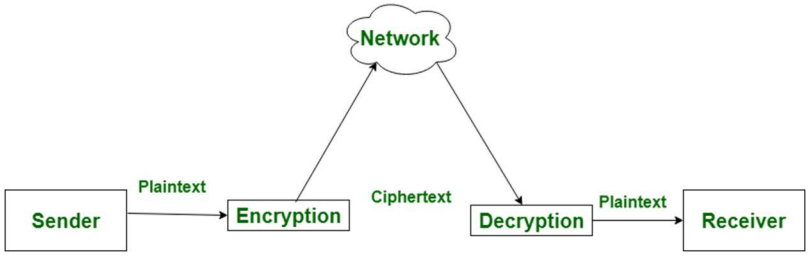
It is rightly claimed that a terrorist with an active internet connection is far more dangerous than otherwise. Small wonder that there is an increase in the incidence of cyber-crimes by terrorists and attempts to breach information systems results in cyberattacks causing colossal losses to ordinary users. They might go about embedding clandestine messages in the form of images through the technique of steganography. It involves concealing information within a text, image or physical object to avoid detection[5]. This information is decoded and extracted at the intended destination. Steganography can be looked in tandem with 'encryption' which is also common as a means of maintaining confidentiality. Increase in the use of encryption is partly due to free use of online encryption. Encryption

[3] M. Anderson, M. Faverio, & J. Gottfried. "Teens, Social Media and Technology," (Pew Research Center, 2023).

[4] Brian Dean. "X (Twitter) Statistics How Many People Use X?" *Backlinko* (23 September 2024), at https://backlinko.com/twitter-users.

[5] Available at: https://www.kaspersky.com/resource-center/definitions/what-is-steganography.

converts a message into an indecipherable format which can only be interpreted by the receiver with the decryption key.



Source: https://www.geeksforgeeks.org/difference-between-encryption-and-decryption/.

For instance, the role of Daesh in recent past demonstrates how social media platforms could turn into 'arsenal' and means of recruiting terrorists, yielding social media platforms into battlegrounds. Social media gets 'weaponized' in the hands of terrorists when cyber terrorism becomes a conscious, deliberate, and premeditated attack on political regimes, governments, information systems, financial institutions, and security institutions.[6]

The kinship between Twitter and political engagement became pertinent with political uprisings in Tunisia and Egypt fueled by public campaigns on Twitter in 2011. Not only in the past but also in recent times contemporary politics is rife with examples where political actors across the globe have utilized Twitter to connect with a larger audience base. This neutralizes effects of heavily censored and mediated communication broadcast on national media. More so, significant patterns of public engagement were discernible during 2010 Australian Federal Elections, likewise in 2012 US Presidential Elections which highlighted enfranchisement of public opinion through social media platforms, particularly Twitter.[7]

The significance of X increased manifold as not only indigenous but international users took to Twitter to access news regarding Ukraine protests in 2014. The number of tweets soared as events turned more volatile. As the events unfolded, the frequency of tweets increased, precipitating into twitter

[6] Joseph Shaheen. How Daesh Uses Adaptive Social Networks, *NATO Report* (November 2015). Visit at https://stratcomcoe.org/cuploads/pfiles/full_report_joe_12-01-2016.pdf.

[7] A. Bruns and J. E. Burgess, "The Use of Twitter Hashtags in the Formation of Ad hoc Publics", in 6th European Consortium for Political Research General Conference, 25 - 27 August 2011, University of Iceland, 2011.

storms led by the hashtag 'Euromaidan'. Not only on Twitter did it peak the trending topics but it also became a record breaker by drawing in more than 200,000 followers on the Euromaidan Facebook page.

Since Twitter use was less pronounced in Ukraine as compared to Facebook, crowd pullers took to Facebook for exchanging information, planning as well as for mass mobilization of requisite resources.[8] This further reflected in the use of language employed by both media. As the use of Facebook has been in common currency, it employed local language while Twitter targeted global audience with the use of English.[9] Twitter storms, trending hashtags, and newsfeeds brought the Ukrainian issue in the limelight earning sympathy and global attention so much so that the American Assistant to the Secretary of State paid a visit to Euromaidan in Kiev.[10] The mounting social media pressure cast a great impact on geopolitics. US Secretary of State, John Kerry, issued a statement in the favour of Ukrainian protestors, denouncing the highhandedness of Ukrainian authorities in quelling peaceful protestors through the use of brute police force. The gruesome scenes of batons and bulldozers was seen as a breach of basic democratic rights of peaceful citizens.[11]

This can be compared with a spike in Twitter activity against the backdrop of Ukraine-Russia war in the aftermath of Russian invasion of February 24, 2022. Accessing dataset of tweets[12] shows a direct public communication and engagement with Russian state-sponsored media.[13] During times of soaring conflict between Russia and Ukraine, information warfare became rampant where Ukrainian people continued to speak for their cause. Likewise, wars are no longer fought on the geographical fronts but also occupy the virtual space. Twitter's streaming API garnered trending hashtags and keywords. The major fifteen hashtags were as under:

---

[8] Christopher M. Buchanan. "Social Media's Role in Shaping Warfare" (Master's thesis, United States Marine Corps and Staff College, 2017).

[9] Ibid.

[10] Megan Metzger, Pablo Barbera, and Penfold-Brown. "Ukraine Protests 2013-2014". *Social Media and Political Participation Lab* (Report, NY University, February 2014).

[11] "Top U.S. Official Visits Protesters in Kiev as Obama Admin. Ups Pressure on Ukraine President Yanukovich," *CBS New*, December 11, 2013, at http://www.cbsnews.com/news/us-victorianuland-wades-into-ukraine-turmoil-over-yanukovich/, (accessed September 23, 2024), 1

[12] This dataset can be found at https://github.com/echen102/ukraine-russia.

[13] Ibid.

#ukraine, #russia, #putin, #standwithukraine, #ukrainerussiawar, #nato, ##russian, #ukrainian, #kyiv, #Ukrainianwar, #zelensky, #Mariupol, #stoprussia, #slavukraini, #tigray.[14]

The recurrence of these hashtags waned in the coming months and other hashtags like #StopPutin and #SafeAirliftUkraine gained currency. The hashtag #putin, however, remained on top with only #zelensky corresponding to it in terms of popularity or even surpassing it. There were a number of tweets which addressed the US president Joe Biden to help salvage the hapless Ukrainian citizens. Contrastingly, the hashtags using names of American Presidents such as Trump and Biden could not match the frequency with which Putin had been addressed. There was an upsurge in use of hashtags mentioning Putin especially in March 2022 also because Russia did not comply with agreements despite deliberations on temporary ceasefires.[15]

Pakistan Experience

Twitter is one of the leading data sources on the international arena influencing socio-political milieu. Twitter datasets are often used in the domains of computational politics to assess levels of polarization and propaganda alongside political trolling. It is interesting to note that in a country like Pakistan where political climate is marred by dynastic politics, threats of religious extremism, ethnolinguistic conflicts, and with a concurrent low rate of internet penetration as far back as in 2013 and underdeveloped telecommunication infrastructure, the 2013 General Election turned out to be the country's first 'youth election'. Analysing the data for voter age bracket authenticated the claim. The data demonstrated that 35 per cent of the total voter turnout on the national level comprised voters under 30 years of age, while 20 percent of the voters were in the age bracket of 18-25.[16] When scaled at the provincial level, it turned out that 24 per cent of eligible voter population comprised youth aged (18-25) in KPK.

---

[14] Emily Chen and Emilio Ferrara. "Tweets in Time of Conflict: A Public Dataset Tracking the Twitter Discourse on the War Between Ukraine and Russia," *Proceedings of the International AAAI Conference on Web and Social Media* 17 No. 1 (2023):1006-13. Visit https://doi.org/10.1609/ icwsm.v17i1.22208.

[15] Ibid.

[16] Saifuddin Ahmed, Marko M. Skoric. "My Name is Khan: The Use of Twitter in the Campaign for 2013 Pakistan General Election", *47th Hawaii International Conference on System Science* (Hawaii, 2014). https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6758880.

This was a major difference with respect to other provinces such as Punjab and Sindh in 2013.[17]

The 2018 General Elections of Pakistan saw a colossal increase in the use of social media campaigns catapulting parties with marginal status to evolving into a formidable political power to reckon with due to their digital acumen in handling social media platforms particularly Twitter. Through a careful content analysis of available tweets from top four political parties, studies reveal that only one of the parties was able to garner most public support owing to public interaction at close quarters, keeping the general public updated with campaign dates and whereabouts and large scale public mobilization to caste vote. The inclusion of women and particularly youngsters was a distinct feature hitherto conspicuously absent in political canvassing in Pakistan.[18] Thus, Twitter communication coupled with public campaigns led to increased voter turnout especially among youth.

Pakistan witnessed a spike in internet users with 111.0 million new internet users from the very onset of 2024 coinciding with national elections which were held on February 8, 2024. The outreach of internet has penetrated 45.7 percent of the population.[19] Statistics demonstrate highest number for YouTube users with 71.7 million active users, followed by Facebook with 57.5 million users. Twitter with its 3.4 million users might appear meagre but its impact has been monumental on current socio-political situation. This was further vouched by PTA declaring 8 top social media platforms.[20]

In April 2022, Pakistan underwent a political upheaval following the no-confidence motion against the then Prime Minister. This led to a large scale twitter storm with trends denouncing this move, demanding earlier election, while on the other hand drew ire equally from other parties favouring this move.

---

[17] Ibid.

[18] M. Furqan Rao. "Hate Speech and Media Information Literacy in the Digital Age: A Case Study of 2018 Elections in Pakistan," *Global Media Journal* 18, No.34 (March 2020). At https://www. globalmediajournal.com/open-access/hate-speech-and-media-information-literacy-in-the-digital-age-a-case-study-of-2018-elections-in-pakistan.php? aid=87925.

[19] Visit https://datareportal.com/reports/digital-2024-pakistan.

[20] "PTA Reveals Top 8 Social Media Platforms of Pakistan," at https://propakistani.pk/2023/01/ 12/pta-reveals-top-8-social-media-platforms-of-pakistan/.

The Twitter API provides location of the users on account of profiles. As discussed earlier, anonymity pervades online forums with respect to sensitive content and more so in tandem with exchange of insights on political issues. A close scrutiny of the API revealed that only a small percentage of data yielded information regarding geographical location of Twitter users. However, based on the shared information, it turned out that tweets flooded from approximately 129 countries other than Pakistan. Most of the tweets came from US, UK, UAE, and Saudi Arabia as these countries host a large section of immigrant population from Pakistan. The data procured from the dataset can help further studies measuring political bias of media outlets, journalists, and other actors involved in political manoeuvring in Pakistan.[21]

In the past two years, a queer mix of online and offline campaigns has emerged as the new defining feature of Pakistani politics with preponderant online engagement, tapping on the wave of change against corruption. This has led to strict disciplinary action on the part of state institutions against those considered as breaching national stability through dissident voices. Hence, X remains banned in the country.

The Role of IoT and AI

The Internet of Things (IoT) has also revolutionized communication incredibly by connecting a vast array of devices and enabling transmission of data through sensors and software with little human intervention. This has further complicated how data and content can be shared through social media with innumerable people accessing shared information. Lack of adequate safety also plays in the interest of miscreants who find it easy to breach media devices, vehicles, baby monitors, and even patients' pacemakers, making patients vulnerable to electric shocks or loss of battery. The overwhelming wireless networks blanketing the global milieu create an atmosphere where anything connected with the internet becomes a means of sending communicable information. IoT has thereby increased security challenges a great deal where social media platforms act as a catalyst to transmit these. The recent pager blasts in Lebanon are a classic example of how IoT can be used or even 'abused' for belligerent purposes.

[21] Visit at https://arxiv.org/pdf/2301.06316.

The predominance of artificial intelligence (AI) in moderating online platforms has sparked considerable debate, particularly concerning its ability to regulate inappropriate behaviour without infringing on users' freedom of expression. There are competing arguments whether such regulating practices thwart human freedom or not. In this context, AI sentiment analysis becomes a viable tool which utilizes natural language processing (NLP) to assess the emotional tone of text and identify potentially harmful content. However, this technological intervention raises significant questions about human freedom vis-à-vis various theories on liberty and freedom of expression. The European Union's regulatory efforts, including the Artificial Intelligence Act and the Digital Services Act, seek to balance the potential benefits of AI with the protection of individual rights. By combining AI with human oversight, transparency, and bias audits, it is possible to develop a moderation framework that both protects users from harm and preserves the core values of freedom and expression in digital spaces. The challenge lies in finding the right balance between surveillance and freedom, ensuring that technological solutions to online behaviour do not inadvertently restrict the very freedoms they are designed to protect.

The European Union's Role in Regulating Social Media

The European Union (EU) has positioned itself as a global standard-bearer in regulating social media, highlighting the need for ethical governance, transparency, and accountability in the digital age. Social media platforms, including Twitter, wield enormous influence over public discourse, information dissemination, and individual privacy. Recognizing the transformative power of these platforms, the EU has implemented comprehensive regulatory frameworks to address systemic risks, safeguard user rights, and ensure a fair digital ecosystem.

More specifically, the right to protection of one's personal date is enshrined in the EU Charter of Fundamental Rights, which always valued 'respect for private life and communication'.[22]

[22] Article 7 & 8, EU Charter of Fundamental Rights at https://fra.europa.eu/en/eu-charter/article/8-protection-personal-data?page=1.

General Data Protection Regulation (GDPR)

The EU adopted the General Data Protection Regulation (GDPR) in 2016, which is a cornerstone of Data Protection.[23] The GDPR regulations has been effective since May 2018, is a foundational pillar of the EU's digital regulation strategy. These set of the EU's regulations is applicable on all companies which process users' data within the ambit of the EU. It establishes stringent requirements for data collection, processing, and storage, mandating social media companies to obtain explicit user consent before handling personal information. The regulation grants users the "right to be forgotten," allowing them to request the deletion of their data. Non-compliance with GDPR can result in significant financial penalties, as exemplified by fines imposed on major tech firms like Meta for privacy violations. This landmark regulation has set a global benchmark for data protection and privacy, influencing policies beyond the EU's borders.[24] Broadly speaking, the rights that GDPR grants can be summed up as under:

i. The right to access, revise or erase personal data referred to as 'the right to be forgotten'.
ii. The right to receive one's personal data and transmit to another company which is known as 'data portability'.
iii. Receive notification in case of any kind of violation of personal data.[25]

In the 27 EU countries, the GDPR regulations have been implemented. As per the EU's unique model, the execution of the GDPR is done independent of government intrusion so that the rules apply across the EU countries through European Data Protection Board. The case of Irish Data Protection Commission imposing a huge fine of €1.2 billion on Meta owing to breach of GDPR rules is a classic example of the EU's strict measures in dealing with the cases of breach of personal data.[26]

---

[23] European Commission, "General Data Protection Regulation (GDPR)," at https://ec.europa.eu/info/law/law-topic/data-protection_en (accessed May 15, 2024).

[24] European Commission, "Code of Practice on Disinformation," at https://digital-strategy.ec.europa.eu/en/ policies/online-disinformation (accessed October 1, 2023).

[25] EPRS, "Regulating Social Media: What is the European Union Doing to Protect Social Media Users," (June 2024), at https://ep thinktank.eu/2024/06/28/regulating-social-media-what-is-the-european-union-doing-to-protect-social-media-users/.

[26] "Data Protection Commission announces conclusion of inquiry into Meta Ireland", *Data Protection Commission*, 22 May 2023, at https://www.dataprotection.ie/en/news-media/press-releases/Data-Protection-Commission-announces-conclusion-of-inquiry-into-Meta-Ireland.

The EU Digital rulebook

Since 2002, the EU introduced two major laws to ensure safe online environment. The underlying principle behind the two laws i.e. the Digital Markets Act and Digital Services Act was that whatever was illegal offline, should be deemed illegal online.[27]

The Digital Markets Act

The Digital Markets Act entails the creation of a level playing field for all digital companies, allowing start-ups and small firms to compete with bigger industrial units. The European Commission has designated clear rules, especially for large platforms or 'gatekeepers' to keep them from levying unfair conditions on businesses and clientele. In this regard, Alphabet (Google, YouTube), Apple, Byte Dance (TikTok), Meta (Facebook, Instagram and WhatsApp), and Microsoft have been targeted as 'gatekeepers'. As per Digital Markets Act, these platforms will not tilt in the favour of their own services and products over those offered by other parties over their platforms.[28]

The European Commission is the only authority vested with the responsibility of enforcing the regulation, backed by an advisory committee to facilitate its work. Once a large online company is designated as a 'gatekeeper', compliance with the rules of the regulation within 6 months becomes mandatory. In case of a breach on the part of a gatekeeper, it risks:

- a fine approximately 10% of its total worldwide turnover;
- another penalty of approximately 20% of its worldwide turnover for recurrent offences;
- periodic penalty payments of almost 5% of its average daily turnover;
- non-financial structural remedies like selling-off of (parts of) its business, as a last resort for systematic failure to comply.

The Digital Services Act: Addressing Systemic Risks

The EU's Digital Services Act (DSA) adopted in 2022, marks a paradigm shift in regulating online platforms. [29] It imposes obligations on platforms such as Twitter to assess and mitigate systemic risks, including the spread of illegal

[27] EPRS, "Regulating Social Media".

[28] https://eur-lex.europa.eu/EN/legal-content/summary/digital-markets-act.html.

[29] European Commission, "Digital Services Act (DSA)," https://digital-strategy.ec.europa.eu/en/policies/digital-services-act (accessed January 31, 2024).

content, misinformation, and harmful speech. Platforms must implement robust content moderation policies, conduct independent audits, and provide transparency reports detailing their efforts to curb harmful activities. Additionally, the DSA introduces mechanisms for users to appeal content removal decisions, promoting accountability and fairness in content regulation.

Combating Misinformation and Hate Speech

Together with DSA, the EU's broader Gender Equality Strategy (2020-2025) accentuates the need to tackle online gender-based violence and ensure women's safety in the digital environs. The strategy includes provisions for online reporting systems and creating safer digital spaces for women. This means that Twitter, as a platform operating within the EU, must align with the EU's commitment to mitigate offences against women in online environments. The strategy encourages platforms to implement stringent measures to identify and address misogynistic behaviour and harassment, making it a priority for social media companies to create safer spaces for women online.[30]

Algorithmic Transparency and Fairness

Algorithms play a pivotal role in shaping user experiences on social media, determining what the content users see and interact with. However, these systems can unintentionally amplify biases, misinformation, or harmful content. The EU's regulatory frameworks emphasize algorithmic transparency, requiring platforms to disclose how algorithms operate and provide users with the ability to opt-out of personalized recommendations. This focus on fairness and accountability aims to prevent manipulation and foster trust in digital platforms.[31]

The DSA also introduces a requirement for platforms to conduct risk assessments to identify and mitigate the potential for harm. Twitter must assess the risks its platform poses, particularly regarding the amplification of hate speech and harmful content targeting women. If the platform is found to have vulnerabilities that allow such behaviour to thrive, it is required to

[30] Gender Equality Strategy (2020-2025). Visit https://ec.europa.eu/newsroom/just/items/682425/en.

[31] M. Cappello (ed.), Algorithmic transparency and accountability of digital services, *IRIS Special* (Strasbourg: European Audiovisual Observatory, 2023). https://rm.coe.int/iris-special-2023-02 en/1680aeda48.

implement measures to address these risks. This could involve changing the algorithms that promote harmful content or ensuring that accounts engaging in abusive behaviour are swiftly penalized or banned. By focusing on both individual content and systemic factors such as algorithms, the DSA aims to reduce the prevalence of online trolling, which has been particularly pervasive against women, particularly public figures, activists, and journalists.

The General Data Protection Regulation (GDPR) also plays an important role in safeguarding women on social media platforms. While the GDPR primarily focuses on data protection, it also includes provisions related to the protection of vulnerable groups, including women. Under the GDPR, Twitter is required to be transparent about how it collects and uses user data, ensuring that the platform does not use personal information to exacerbate online harassment. This regulation also empowers users, including women, to have control over their personal data and to request that harmful or inaccurate information be removed. This is particularly relevant in cases where women are targeted by trolls who spread false or malicious content about them.

Moreover, the EU has increasingly held platforms like Twitter accountable for failing to prevent online harassment. The EU has repeatedly warned social media platforms, including Twitter, that failure to implement these protective measures could result in significant penalties.

Global Implications of EU Regulations

The EU's proactive approach to social media regulation has far-reaching implications, influencing global digital governance. Major platforms operating within the EU must adapt their practices to comply with its regulations, often extending these changes to users worldwide. This extraterritorial impact underscores the EU's role as a pioneer in setting standards for digital responsibility. Scholars argue that the EU's regulatory frameworks serve as a blueprint for other regions grappling with similar challenges.[32]

[32] Anu Bradford. "The European Rights-Driven Regulatory Model" (3rd chap.) in *Digital Empires: The Global Battle to Regulate Technology* (OUP, 2023).

Challenges and the Road Ahead

Despite its leadership, the EU faces challenges in enforcing its regulations effectively. The dynamic nature of social media technologies, coupled with resistance from platform operators, requires constant vigilance and adaptation. The EU's focus on collaboration with international partners, civil society, and industry stakeholders will be critical in addressing emerging risks and ensuring that regulatory frameworks remain robust and relevant.

The DSA represents a significant regulatory framework for social media platforms operating within the EU. The DSA imposes strict requirements on platforms like Twitter, particularly regarding the moderation of illegal content. Under this law, social media platforms must take proactive steps to identify and remove illegal content, including hate speech, children related abusive material, and terrorist propaganda, while also ensuring the freedom of expression remains protected. One of the most notable provisions of the DSA is the obligation for very large online platforms (VLOPs) to conduct risk assessments concerning the dissemination of harmful content, disinformation, and the use of their platforms for illegal activities.[33] Twitter, being one of the largest social media platforms, is subject to these requirements and must comply with the DSA's transparency and accountability rules.

EU's Perspective on Twitter

From the EU's perspective, social media platforms are integral to democratic discourse but must adhere to stringent standards of transparency, accountability, and data protection. The EU also emphasizes on the responsibility of platforms like Twitter to combat illegal content, misinformation, and hate speech while respecting users' fundamental rights. Moreover, Twitter's role in shaping political narratives has raised concerns about algorithmic bias and echo chambers, prompting the EU to advocate for algorithmic transparency. The General Data Protection Regulation (GDPR) further underscores the EU's commitment to safeguarding user privacy, requiring platforms to manage data responsibly. As it is more evident that Twitter has become a more useful source of promoting far-right populist narratives across Europe, particularly, during national and the EU

[33] European Commission. DSA: Very Large Online Platforms and Search Engines, at https://digital-strategy.ec.europa.eu/en/policies/dsa-vlops.

parliamentary elections.[34] Thus, concerns raised in euro-centric Union is a logical outcome. The European Commission has suspended advertising the EU on X in November 2023 over alleged increase in disinformation regarding the EU policies and hate speech on the platform.[35]

As Twitter evolves under new ownership and introduces policy changes, the EU remains vigilant, stressing compliance with its digital regulations. Balancing free expression with moderation of ethical content is a persistent challenge. Thus, Twitter's trajectory in Europe illustrates the broader tension between innovation and regulation in the digital age, shaping the platform's role in modern society.[36]

Thus, it can be said that the European Union's comprehensive approach to regulating social media reflects its commitment to fostering a safe, inclusive, and transparent digital environment. Through initiatives like the GDPR and DSA, the EU has not only safeguarded the rights of its citizens but also set a global precedent for ethical digital governance. As social media continues to evolve, the EU's regulatory efforts will play a pivotal role in shaping the future of the digital landscape.

Conclusion

Thus, it can be concluded that the immediacy and outreach of X results in rapid dissemination of information, making it a powerful tool for both advocacy and manipulation. Political actors and interest groups have learned to exploit the platform's features, using targeted messages to sway public opinion, mobilize supporters, and undermine opponents. The ability to craft narratives that resonate emotionally with users can lead to the viral dissemination of misinformation and disinformation, blurring the lines between fact and fiction.

---

[34] L. Alonso-Muñoz and A. Casero-Ripollés. "Populism Against Europe in Social Media: The Eurosceptic Discourse on Twitter in Spain, Italy, France, and United Kingdom During the Campaign of the 2019 European Parliament Election," *Frontier Communication* 5, No. 54 (2020).

[35] https://www.euronews.com/next/2023/11/17/eu-commission-advises-services-to-stop-advertising-on-elon-musks-x.

[36] "Twitter will respect EU laws to combat disinformation, Elon Musk says", *Euronews*, 20 June 2023.