Volume 8, 2022, 139-144 https://doi.org/10.5281/zenodo.

TAKING OVER TWITTER – BALANCING FREE SPEECH AND CONTENT MODERATION

Dan Valeriu VOINEA, Ph.D., University of Craiova, Romania

Abstract

In this paper we examine the significant changes in content moderation and free speech policies at Twitter following Elon Musk's controversial $44 billion acquisition in 2022. We trace the takeover process and subsequent organizational shifts, including massive layoffs and altered content guidelines. The study explores the complex balance between fostering free expression and mitigating harmful content on social media platforms, contextualizing Twitter's challenges within broader debates on digital governance and online discourse. The research analyzes Twitter's pre- and post-acquisition approaches to content moderation, highlighting Musk's "free speech absolutist" stance and its implications. Key policy changes are discussed, including reduced content removal, the "Freedom of Speech, not Freedom of Reach" principle, and expansion of user-driven moderation tools like Community Notes. The paper also addresses concerns raised by critics, advertisers, and governments regarding potential increases in misinformation and harassment under the new regime. Ultimately, this analysis underscores the ongoing tensions between free speech advocacy and responsible platform management in the digital age. It offers insights into the evolving landscape of social media governance and its wider societal impacts, particularly in light of Musk's influential role at Twitter.

Keywords: *Content moderation; Twitter acquisition; Elon Musk; Free speech; Social media governance*

Introduction

The saga of Elon Musk's Twitter acquisition began in early 2022 when the Tesla and SpaceX CEO revealed a 9.2% stake in the social media platform, making him its largest shareholder. (ABC NEWS, 2022) This move set the stage for one of the most dramatic and controversial takeovers in recent corporate history. On April 14, 2022, Musk made an unsolicited bid to buy Twitter for $54.20 per share, valuing the company at a staggering $44 billion. The Twitter board's initial response was to adopt a "poison pill" strategy, but they eventually accepted Musk's offer on April 25. (Conger & Hirsch, 2022) However, the deal was far from smooth sailing. Musk attempted to back out, citing concerns about the number of bot accounts on the platform, which led Twitter to sue him in the Delaware Chancery Court to force the completion of the acquisition. After months of legal battles and public controversies, Musk finally closed the deal on October 27, 2022. (Fung, 2022) His first act as the new owner was to fire top executives, including CEO Parag Agrawal, signaling the beginning of significant changes at the company. (Thomas & Corse, 2022)

In the months following the acquisition, Musk implemented sweeping changes at Twitter. He dramatically reduced the workforce, cutting approximately 50% of the staff. (Botros, 2022) Content moderation policies were altered, and Musk introduced Twitter Blue, a subscription service offering verification badges. The takeover and subsequent changes sparked numerous controversies. Critics raised concerns about the potential spread of misinformation and hate speech on the platform under Musk's leadership. Many advertisers fled the platform, citing worries about content moderation. The acquisition also ignited intense debates about free speech, platform governance, and the power of tech billionaires.

Musk's acquisition of Twitter represents a watershed moment in social media history. Its ramifications extend beyond the platform itself, touching on issues of digital public squares, content moderation, and the role of social media in society. The long-term impact of this takeover on and the broader social media landscape continues to be a subject of ongoing analysis and debate. As with any major corporate transformation, the full consequences of Musk's Twitter acquisition may take years to fully manifest. What's certain is that this takeover has reshaped one of the world's most influential social media platforms and sparked crucial conversations about the future of online communication and in this paper we will try to look at how free speech and content moderation were, are and will be balanced at Twitter.

Free speech, moderation and Twitter

Balancing free speech and content moderation on social media platforms like Twitter involves navigating complex ethical and legal landscapes. Twitter has historically prided itself on being a platform that supports free speech, famously describing itself as "the free speech wing of the free speech party" (Loe, 2022). However, as the platform grew, the need for content moderation became inevitable due to the sheer volume and diversity of user-generated content.

Content moderation involves removing or restricting posts that violate the platform's guidelines, which can include hate speech, misinformation, and harmful content. Twitter, like other platforms, uses a combination of automated systems and human moderators to manage this process. Automated systems are the first line of defense, identifying and flagging potential violations, while human moderators handle more nuanced cases (Loe, 2022).

The challenge lies in striking a balance between allowing free expression and preventing harm. This balancing act is further complicated by varying definitions of harmful content across different regions and cultures. For instance, what might be considered hate speech in one country could be seen as permissible speech in another. This global variation necessitates a flexible and adaptive approach to moderation (Westling, 2022).

Moreover, recent legal and regulatory pressures have added layers of complexity. In the United States, discussions around Section 230 of the Communications Decency Act (Douek, 2022), which provides immunity to online platforms for user-generated content, highlight the ongoing debate about the extent of platforms' responsibilities. Changes to Section 230 could significantly impact how platforms like Twitter moderate content, potentially making them more liable for user posts and possibly leading to either overly aggressive moderation or insufficient content control (Westling, 2022).

In the European Union, the Digital Services Act aims to create a more transparent and accountable framework for online platforms, requiring them to enhance their content moderation practices while protecting users' fundamental rights (Loe, 2022).

Social media platforms, including Twitter, are crucial for fostering discourse, civic engagement, and democratic ideals (Ledesma-Gumasing & Mendoza-Armiendo, 2021). However, the legitimacy and inclusivity of these platforms are vital for maintaining democratic practices (Ledesma-Gumasing & Mendoza-Armiendo, 2021). Despite aiming for inclusivity, Twitter often showcases homogeneous views, particularly from extreme political factions (Zhou et al., 2019). This underscores the challenge of balancing free speech with preventing the dissemination of harmful or extremist content.

Twitter faces a complex challenge in balancing free speech with content moderation. The platform has been criticized for its permissive free speech rhetoric and ineffective enforcement mechanisms (Konikoff, 2021). This tension is exacerbated by the necessity to combat abuse, hate speech, and toxicity to improve user experience (Konikoff, 2021). The platform must revise its policies to detoxify the environment while respecting free speech principles.In the digital era, social media platforms are increasingly involved in regulating free speech, as evidenced by conflicts such as the Nigerian government's clash with Twitter over content removal (Ayalew, 2021). Effective content moderation on Twitter is essential to strike a balance that enhances user experience without overly restricting speech (Durán, 2022). The economics of content moderation on platforms like Twitter involve theoretical considerations and experimental evaluations to achieve this balance (Durán, 2022). Furthermore, the presence of bots on Twitter and other social networks adds complexity to content dynamics on the platform (Schuchard et al., 2019). While retweeting practices can lead to diverse conversational outcomes, intentional retweeting is often aimed at fostering discussions or sharing information (Schuchard et al., 2019). This underscores the intricate relationship between user behavior, platform design, and content dissemination on Twitter.

Elon Musk's acquisition of Twitter has brought new challenges and changes. Musk declares to advocate.for a more open approach to free speech on the platform, yet this vision must be reconciled with the need to manage misinformation and harmful content effectively. His tenure has already seen significant shifts in policy and practice, reflecting the ongoing tension between promoting free expression and ensuring a safe online environment (Westling, 2022).

Twitter's endeavor to balance free speech and content moderation is a multifaceted challenge that necessitates ongoing evaluation and policy adaptation. Achieving the right equilibrium is crucial for cultivating a healthy online environment while upholding the principles of free expression.

Post takeover changes

In the pre-Musk era, Twitter had established a set of community guidelines and employed a combination of algorithmic and human moderation. The platform focused on removing content that violated its rules, which included hate speech, harassment, violent threats, and misinformation, particularly regarding elections and public health. Twitter also introduced labels for potentially misleading information and implemented a strike system for repeat offenders. However, the landscape changed dramatically following Elon Musk's takeover. Musk, who proclaimed himself a "free speech absolutist," promised to transform Twitter into a bastion of free expression. This vision led to several significant changes in the platform's approach to content moderation.

Elon Musk's acquisition of Twitter has ignited discussions and concerns about the platform's approach to free speech and content moderation. Musk, known for advocating for free speech, has expressed his intention to steer Twitter in a new direction, potentially allowing previously banned accounts to be reinstated (Stokel-Walker, 2022). However, this move has also raised apprehensions about the platform's vulnerability to misinformation and disinformation. ("Musk's plans for Twitter to emerge by year-end", 2022) Musk has emphasized the importance of free speech while ensuring compliance with local laws, including content screening regulations ("Musk's Twitter takeover highlights disinformation risk", 2022). The acquisition has intensified Twitter's challenges, with implications for issues such as online radicalization and the platform's societal role (Kouba, 2022). Furthermore, the potential impact on disinformation and the increased focus on misinformation due to Musk's involvement underscore the intricate balance Twitter must strike between free speech and combating harmful content (Sparkes, 2022).

Twitter also significantly reduced its trust and safety team, signaling a shift away from aggressive content moderation. Instead of removing certain types of content, the platform began to focus on reducing the visibility of controversial posts, adopting a policy that Musk described as "Freedom of Speech, not Freedom of Reach." The platform also expanded its Community Notes feature, allowing users to add context to potentially misleading tweets. This move aligned with Musk's stated goal of increasing transparency in content moderation decisions, which included the release of the "Twitter Files."

The acquisition by Musk has led to discussions about the impact on contentious actors on Twitter (Barrie, 2022). The presence of such actors on the platform raises concerns about the potential amplification of divisive or harmful content under Musk's leadership. Addressing the behavior of contentious actors and ensuring a healthy discourse environment will be crucial for Twitter's future under Musk's ownership.

Elon Musk's influence on Twitter post-acquisition highlights the complex interplay between advocating for free speech, content moderation, and the platform's societal influence. As Twitter manages this transition, it will be crucial to monitor how Musk's vision for the platform aligns with principles of free expression while addressing concerns related to misinformation, radicalization, and online discourse.

Despite these changes, Twitter continues to grapple with the challenges of balancing free speech and content moderation. The platform has faced criticism from various quarters. Many advertisers left due to concerns about brand safety in a less moderated environment. Governments worldwide have increased their scrutiny of social media content moderation practices. Critics argue that the looser moderation policies have led to an increased spread of misinformation and concerns persist about the platform's ability to protect vulnerable users from harassment and hate speech.

After Elon Musk's acquisition of Twitter, significant changes were observed in the platform's moderation policies. Musk's takeover led to a strategic realignment of Twitter, with a focus on capturing research results systematically to facilitate future comparative longitudinal studies of usage changes on the platform (Pilgrim & Bohnet-Joschko, 2022). One notable change was the dissolution of Twitter's Trust and Safety council, which was responsible for addressing issues like hate speech and harmful content (Fitzgerald, 2022). This move signaled a shift towards looser content moderation standards aligned with Musk's advocacy for free speech absolutism.

In response to growing concerns about misinformation, Twitter implemented changes to its content moderation policies during the COVID-19 pandemic (Calac et al., 2022). These changes included labeling tweets containing vaccine misinformation, removing misleading content deemed harmful to the public, and suspending accounts posting COVID-19 and vaccine-related misinformation. This proactive approach aimed to combat the spread of false information and ensure the dissemination of accurate and reliable content on the platform, but backfired by enabling a powerful anti-vaccine agenda.

Elon Musk's acquisition of Twitter also brought attention to the issue of fake accounts and social bots on the platform (Zhou et al., 2022). Moreover, Musk's acquisition of Twitter has raised questions about the platform's approach to moderation and censorship. Platforms like Twitter have adopted policies to attach warning labels to posts containing false or disputed information, rather than resorting to direct censorship (Epstein et al., 2022). This approach aims to curb the spread of misinformation while respecting principles of free speech and open dialogue. In late 2022, Twitter revised its content moderation policies, ending enforcement of rules against COVID-19 misinformation. (O'Sullivan, 2022) Changes to the platform's recommendation algorithms reportedly increased the visibility of unverified claims about the conflict in Ukraine. These adjustments also appeared to boost follower counts for media organizations linked to Russia, China, and Iran. (Kann, 2022)

Conclusions

The acquisition of Twitter by Elon Musk represents a watershed moment in the platform's history and the broader landscape of social media governance. This takeover has sparked intense debates about the delicate balance between free speech and content moderation in digital public squares.

Musk's self-proclaimed "free speech absolutist" approach has ushered in significant changes to Twitter's content moderation strategies. The platform has shifted away from content removal towards reducing the visibility of controversial posts. This philosophical pivot has been accompanied by dramatic organizational restructuring, most notably the substantial reduction in workforce, particularly within trust and safety teams. These changes signal a fundamental shift in Twitter's priorities and operational approach.

The expansion of features like Community Notes exemplifies a move towards user-driven moderation, raising important questions about the platform's role in content curation. This shift towards user empowerment, while innovative, also brings to the fore concerns about platform responsibility and the potential for misinformation spread.

The implications of these changes extend far beyond Twitter's digital borders. The platform's evolving policies have far-reaching consequences, influencing public discourse, information dissemination, and even geopolitical dynamics. Moreover, the loosening of moderation policies has led to concerns about brand safety, resulting in a significant exodus of advertisers from the platform.

The Musk era at Twitter serves as a compelling case study in the challenges of managing a global social media platform. It underscores the tension between idealistic visions of free speech and the practical realities of operating in a complex, multinational environment.

As Twitter continues to evolve under Musk's leadership, several key questions remain at the forefront: How will the platform balance its commitment to free speech with the need to protect users from harassment and misinformation? Can Twitter regain advertiser confidence while maintaining its new approach to content moderation? What will be the long-term impact on public discourse and the spread of information? How will regulatory bodies respond to these changes, particularly in light of growing global concerns about social media's influence?

The Twitter experiment under Musk is far from concluded. Its outcomes will likely influence not only the future of the platform but also shape broader discussions about the role of social media in society, the limits of free speech in digital spaces, and the responsibilities of tech leaders in shaping public discourse.

Moving forward, it is crucial to continue monitoring these developments, fostering open dialogue about the implications of platform governance decisions, and working towards solutions that protect free expression while mitigating the harms of unchecked online communication.

The story of Twitter's transformation serves as a powerful reminder of the complex interplay between technology, society, and individual rights in the digital age. It underscores the need for ongoing research, thoughtful policy-making, and public engagement to navigate the challenges of our increasingly connected world. As we observe and analyze these changes, we contribute to the vital conversation about the future of digital communication and its impact on global society.

References

ABC NEWS. (2022, November 11). A timeline of Elon Musk's tumultuous Twitter acquisition. ABC News. https://abcnews.go.com/Business/timeline-elon-musks-tumultuous-twitter-acquisition-attempt/story?id=86611191

Ayalew, Y. E. (2021). From Digital Authoritarianism to Platforms' Leviathan Power: Freedom of Expression in the Digital Age Under Siege in Africa. Mizan Law Review. https://doi.org/10.4314/mlr.v15i2.5

Barrie, C. (2022). Did the Musk Takeover Boost Contentious Actors on Twitter? HKS Misinfo Review. https://doi.org/10.37016/mr-2020-122

Barrie, C. T. (2022). Did the Musk Takeover Boost Contentious Actors on Twitter? https://doi.org/10.48550/arxiv.2212.10646

Barth, S. (2022, November 28). How Musk's Takeover is affecting Twitter's Rules – Digital Society Blog. HIIG. https://www.hiig.de/en/twitter-policies/

Botros, A. (2022, August 11). Twitter manager who oversaw election team says Elon Musk cutting 50% of employees just before the midterms 'certainly doesn't look good.' Fortune. https://fortune.com/2022/11/07/former-election-team-manager-elon-musk-layoffs-midterms-concern/

Conger, K., & Hirsch, L. (2022, October 28). Elon Musk Completes $44 Billion Deal to Own Twitter. The New York Times. https://www.nytimes.com/2022/10/27/technology/elon-musk-twitter-deal-complete.html

Douek, E. (2022, October). Four questions: Evelyn Douek on what Section 230 is and why it is misunderstood. https://news.stanford.edu/stories/2022/10/four-questions-evelyn-douek-section-230-misunderstood

Durán, R. J. (2022). The Economics of Content Moderation: Theory and Experimental Evidence From Hate Speech on Twitter. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4044098

Fitzgerald, J. (2022). Special Issue: The (International) Politics of Content Takedowns: Theory, Practice, Ethics. Policy & Internet. https://doi.org/10.1002/poi3.375

Fung, C. D., Gabrielle Smith, Brian. (2022, October 28). A look back at Elon Musk's first year owning Twitter, in tweets | CNN Business. CNN.

Golder, S. (2024). Social Media Engagement in Health and Climate Change: An Exploratory Analysis of Twitter. Environmental Research Health. https://doi.org/10.1088/2752-5309/ad22ea

Kann, A. (2022, December). State-controlled media experience sudden Twitter gains after unannounced platform policy change—DFRLab.

Konikoff, D. (2021). Gatekeepers of Toxicity: Reconceptualizing Twitter's Abuse and Hate Speech Policies. Policy & Internet. https://doi.org/10.1002/poi3.265

Kouba, T. (2022). Online Radicalization: Twitter Privatization as a Threat to the Modern Society. Politika Nacionalne Bezbednosti. https://doi.org/10.22182/pnb.specijal2022.9

Ledesma-Gumasing, R., & Mendoza-Armiendo, R. (2021). Popular, Accessible, Inclusive: Social Media as an Ideal for Decision-Making in a Democracy. Journal of Public Policy and Administration. https://doi.org/10.11648/j.jppa.20210504.12

Loe, M. (2022, May 22). How content moderation on social media really works. TechHQ.

Mo, M. (2022). Analyzing Twitter Data to Understand Stigmatization of Schizophrenia Before and After Elon Musk. Journal of Student Research. https://doi.org/10.47611/jsrhs.v12i3.4637

Musk Takeover Deepens Twitter's Turmoil. (2022). https://doi.org/10.1108/oxan-es273677

Musk's Plans for Twitter to Emerge by Year-End. (2022). https://doi.org/10.1108/oxan-db268843

Musk's Twitter Takeover Highlights Disinformation Risk. (2022). https://doi.org/10.1108/oxan-db273846

O'Sullivan, D. (2022, November). Twitter is no longer enforcing its Covid misinformation policy | CNN Business. https://edition.cnn.com/2022/11/29/tech/twitter-covid-misinformation-policy/index.html

Park, A. (2022). SEM Analysis of Agreement With Regulating Online Hate Speech: Influences of Victimization, Social Harm Assessment, and Regulatory Effectiveness Assessment. Frontiers in Psychology. https://doi.org/10.3389/fpsyg.2023.1276568

Schuchard, R., Crooks, A., Stefanidis, A., & Croitoru, A. (2019). Bot Stamina: Examining the Influence and Staying Power of Bots in Online Social Networks. Applied Network Science. https://doi.org/10.1007/s41109-019-0164-x

Sparkes, M. (2022). Musk's Twitter Takeover Puts Misinformation in the Spotlight. The New Scientist. https://doi.org/10.1016/s0262-4079(22)02019-x

Thomas, L., & Corse, A. (2022, October 28). Elon Musk Closes Twitter Deal, Immediately Fires Top Executives. Wall Street Journal. https://www.wsj.com/articles/elon-musk-completes-twitter-takeover-11666918031

Westling, J. (n.d.). Lawmakers' Misguided Approach to Social Media Content Moderation. AAF. Retrieved July 24, 2024, from https://www.americanactionforum.org/insight/lawmakers-misguided-approach-to-social-media-content-moderation/

Zhou, Y., Dredze, M., Broniatowski, D. A., & Adler, W. D. (2019). Elites and Foreign Actors Among the Alt-Right: The Gab Social Media Platform. First Monday. https://doi.org/10.5210/fm.v24i9.10062