Jonathan Aaron Rocha

ENGL 5374 – Research Methods in Graduate Studies

Dr. Amber Dunai

December 6, 2024

The Illusion of Absolute Free Speech: How Twitter's Evolution to X Reveals the Paradox of Platform Governance

The rapid evolution of social media platforms as central spaces for public discourse has fundamentally transformed how society approaches free speech and content moderation. As these digital spaces increasingly function as de facto public squares, the policies governing user expression have become crucial determinants of democratic discourse. Twitter's transformation to X under Elon Musk's ownership presents a unique case study of how platform governance affects public communication. This transition, marked by dramatic policy shifts and public declarations about free speech absolutism, offers an opportunity to examine the complex relationship between platform ownership, content moderation, and user agency.

Recent scholarship has highlighted the growing tension between platforms' roles as speech facilitators and their responsibilities as content moderators. Evelyn Douek's analysis of content moderation evolution suggests that platforms must move beyond simplistic approaches to speech governance (Douek 12). At the same time, Tarleton Gillespie argues that content moderation is fundamental to platform identity and function. Meanwhile, studies like Wang et al.'s examination of user responses to Musk's takeover reveal the complex dynamics between user expectations and corporate control in digital spaces (Wang et al. 8).

This essay argues that Twitter's rebranding as X under Elon Musk highlights the paradox of free speech absolutism, where promises of unrestricted expression mask more profound corporate control and user disempowerment. By comparing Twitter's pre- and post-acquisition policies, the essay demonstrates how changes in platform governance reflect broader tensions between free speech ideals and the practical challenges of content moderation in digital spaces. It also argues that the platform's policy changes have prioritized corporate interests over democratic discourse and user empowerment.

The scholarly discourse surrounding platform governance and content moderation encompasses legal frameworks, sociocultural analyses, and studies of user agency. Understanding these perspectives provides crucial context for analyzing Twitter's policy evolution under Musk's ownership.

Evelyn Douek's content moderation analysis provides a foundational framework for understanding how platforms approach speech governance. In "Governing Online Speech: From 'Posts-as-Trumps' to Proportionality and Probability," Douek argues that platforms must move beyond simplistic, case-by-case content decisions toward more systematic approaches that balance free speech with harm prevention (Douek 12). This shift reflects the growing recognition that content moderation requires nuanced frameworks rather than absolute rules.

Kate Klonick's concept of platforms as "new governors" further illuminates the quasi-governmental role that social media companies play in regulating online speech. In "The New Governors: The People, Rules, and Processes Governing Online Speech," Klonick demonstrates how platforms have developed complex systems of governance that operate outside traditional legal frameworks while significantly impacting public discourse (Klonick 1627). Klonick's framework of platforms as 'new governors' becomes especially relevant when examining X's

governance evolution (Klonick 1627). The platform's transition exemplifies what Klonick identifies as the tension between private corporate interests and public governance responsibilities. Under Musk's ownership, this tension manifests in policies claiming to expand speech rights while consolidating corporate control over content visibility and user reach. This transformation provides a concrete example of how platform governance can shift from what Klonick terms' democratic-adjacent' systems toward more centralized corporate control structures. This perspective is particularly relevant when examining how ownership changes affect platform governance structures.

Tarleton Gillespie's analysis in *Custodians of the Internet* reveals how content moderation is not merely an ancillary function but central to platform identity and operation (Gillespie 23). Gillespie argues that platforms actively shape public discourse through moderation policies, often prioritizing corporate interests over user agency. His work highlights the tension between platforms' public presentation as neutral spaces and their active role in governing speech.

The societal implications of platform governance extend beyond individual content decisions. Sarah T. Roberts's *Behind the Screen* exposes the human infrastructure of content moderation, revealing how platforms rely on hidden labor to maintain their public spaces. This perspective adds crucial context to understanding how policy changes affect users and the workforce responsible for implementing moderation decisions.

Recent scholarship has increasingly focused on the relationship between user empowerment and platform ownership. Rui Wang et al.'s analysis of Twitter users' responses to Musk's takeover provides valuable insights into how ownership changes affect user perceptions and behavior. Their research reveals the complex dynamics between users' desire for free expression and their concerns about platform stability and responsibility.

Jameel Jaffer's *The Fight for Privacy* (2022) connects platform governance to broader user privacy and autonomy issues. Jaffer argues that content moderation decisions increasingly affect what users can say and how their data and privacy are protected, adding another crucial dimension to discussions of user agency in digital spaces.

David Kaye's *Speech Police: The Global Struggle to Govern the Internet* offers essential insights into how platform governance affects global communication. Kaye's international analysis of different regulatory approaches helps contextualize Twitter's policy shifts within broader international debates about free speech and content moderation.

Safiya Umoja Noble's work on algorithmic bias in *Algorithms of Oppression* demonstrates how platform policies can reinforce existing power structures while appearing neutral (Noble 162). Building on Noble's work, Ruha Benjamin's *Race After Technology* (2019) examines how platform governance decisions often disproportionately impact marginalized communities, mainly when platforms prioritize unrestricted speech over protecting vulnerable users (Benjamin 47).

Jean Burgess and Nancy K. Baym's *Twitter: A Biography* (2020) provides a historical context for understanding Twitter's evolution as a platform. Their analysis of how Twitter's identity has shifted over time helps frame current changes under Musk's ownership within the platform's more extended history of governance changes.

This study employs critical discourse analysis to examine Twitter's content moderation policies before and after Musk's acquisition. This study applies Fairclough's critical discourse analysis framework (1995), emphasizing how institutional language reflects and reinforces power relationships (Fairclough 81). Fairclough's three-dimensional model—examining the text, discursive practice, and social practice—systematically analyzes how Twitter's policy documents

reveal shifting power dynamics between platform owners and users. This framework is particularly relevant for understanding how X's new policies while using the language of free speech, establish different forms of institutional control through linguistic and structural changes in platform governance documentation. Particularly relevant is how changes in policy language reveal evolving relationships between platform owners, users, and content moderators.

The primary texts selected for analysis include Twitter's official policy documents from two distinct periods: pre-acquisition (2021-2022) and post-acquisition (2022-2024). These documents include:

Primary Texts Selected for Analysis of Twitter's Official Policy Documents	
Category	Documents
Pre-Musk Twitter Documents	• Terms of Service (2021-2022 version) • Content moderation guidelines • Hateful conduct policy • Platform rules and community standards • Official blog posts about content moderation decisions
Post-Musk Twitter/X Documents	• Updated Terms of Service • Revised content moderation policies • Public statements about platform governance • Official announcements regarding policy changes • Platform transparency reports

Table 1

The evolution in policy language is particularly evident in specific changes in terms of service. Pre-Musk Twitter's policy stated: 'We believe everyone should have the power to create and share ideas and information instantly, without barriers' (Twitter ToS, 2021), emphasizing collective empowerment. The 2021 Terms of Service emphasized user responsibilities and platform rights to remove harmful or misleading content, a sharp contrast to the broader emphasis on 'visibility filtering' introduced under Musk's ownership (Twitter, Inc.). In contrast, X's updated policy declares: 'X enables free speech in the digital town square' (X ToS, 2023), shifting emphasis to individual expression while paradoxically centralizing control.

The analysis also incorporates supplementary materials such as Musk's public statements about platform policies, official company communications, and documentation of specific moderation decisions during both periods (Musk, "Interview with X News," 2023).

The comparative analysis focuses on several key dimensions:

1. Policy Language and Tone:

   - Examination of how policy documents frame the relationship between platform and users

   - Analysis of how content moderation is justified and explained

   - Changes to the public representation of platform values and priorities

2. Governance Structures:

   - Shifts in decision-making processes

   - Changes in appeal mechanisms

   - Modifications to enforcement procedures

3. User Rights and Responsibilities:

- Evolution of user empowerment language

- Changes in how user agency is described and facilitated

- Shifts in platform-user power dynamics

While this analysis draws on publicly available policy documents and official communications, several methodological limitations warrant acknowledgment. The opacity of algorithmic decision-making processes limits our comprehension of the practical implementation of content visibility decisions. Access to internal documentation, such as algorithmic training data, moderation guidelines, and decision trees for content filtering, would provide valuable insights into the actual mechanisms of content control. Secondly, direct observation of moderation processes, encompassing the interactions between human moderators and automated technologies and applying policy rules in real-time contexts, could augment our comprehension of platform governance. Third, the lack of access to internal metrics about content removal, appeals, and algorithmic filtering limits our ability to quantify the full impact of policy changes on user expression and engagement. Future research would benefit from collaborative partnerships with platforms that could provide access to these internal processes and data while maintaining necessary privacy and security protocols.

Having established the methodological framework for analyzing Twitter's policy evolution, we can now examine how specific policy changes reflect broader shifts in platform governance philosophy. The pre-Musk era of Twitter's content moderation provides a crucial baseline for understanding subsequent transformations in the platform's approach to governing online speech. This analysis reveals how policy changes under new ownership represent tactical adjustments and fundamental shifts in how the platform conceptualizes its role in public discourse. Twitter's pre-Musk content moderation policies reflected a complex balancing act

between promoting free expression and preventing harm. Twitter's pre-Musk Terms of Service explicitly stated that the platform's purpose was 'to serve the public conversation,' emphasizing collective responsibility.

In contrast, X's updated Terms of Service emphasize 'maximizing free speech,' reflecting a more individualistic approach to platform governance. The 2024 Terms of Service explicitly codify content visibility controls, stating, *'We may limit or terminate the availability of the Services or particular features to you or other users at any time'* (X Corp.). The platform's approach stressed what Gillespie calls 'custodial' governance (Gillespie 23), with content filtering presented as a vital component of platform health. This observation aligns with what Roberts describes as the 'institutionalization of content moderation' (Roberts 156), where platform governance becomes systematized and professionalized. The Terms of Service explicitly acknowledged Twitter's role as a curator of public discourse, stating that "Twitter's purpose is to serve the public conversation." The evolution of X's governance model represents what Gillespie terms a 'paradox of custodial governance' (24), where increased automation and reduced human oversight fundamentally alter the platform's custodial role. While Gillespie argues that content moderation is inherent to platform identity, X's shift toward algorithmic control transforms this relationship, moving from active curation to what might be termed 'automated custodianship.' This transformation aligns with Gillespie's prediction that platforms would increasingly rely on technological solutions to govern speech, though perhaps not in the democratizing way he envisioned. This framing positioned the platform as a steward of digital speech rather than merely a neutral host.

While the 2021 Terms framed content removal as a reactive enforcement measure against specific violations, the 2024 iteration expands corporate discretion with clauses allowing

'enforcement actions for any reason, including commercial inviability' (X Corp.). The

enforcement mechanisms during this period followed a graduated approach. Violations typically

resulted in escalating consequences, from warnings to temporary suspensions before permanent

account removal. This system reflected Douek's concept of proportionality in content

moderation, attempting to match enforcement actions with the severity and frequency of

violations. The platform maintained detailed public-facing documents explaining these

enforcement decisions, emphasizing transparency in governance.

Pre-Musk Twitter's stated principles centered on fostering "healthy conversation" while

acknowledging the platform's responsibility to protect marginalized users. The content

moderation policies explicitly addressed harassment, hate speech, and misinformation through

detailed guidelines and enforcement protocols. These policies reflected what Roberts identifies

as institutionalized content moderation, where platform governance is systematized and

professionalized.

Implementation strategies included:

Pre-Musk Acquisition Terms of Service Principles for Twitter	
Principle	**Implementation Strategy**
Foster healthy conversation	Precise categorization of violation types
Protect marginalized users	Detailed appeal processes
Address harassment, hate speech, and misinformation	Regular transparency reports
Systematize and professionalize content moderation	Collaboration with external researchers and experts

Balance automation with human review	Investment in automated moderation tools

Table 2

The transition to X under Musk's ownership marked a significant shift in content moderation philosophy. The new Terms of Service (X Corp. 2024) emphasized "free speech" over content restriction, reflecting what Wang et al. identify as a tension between absolutist free speech ideals and practical governance needs. Notable changes included:

Post-Musk Acquisition of Twitter Terms of Service Principles for the rebranded X	
Aspect	**Impact on User Interface/User Experience**
Philosophical Shift	Emphasis on 'free speech' over content restriction
Terms of Service	Revised to reflect a tension between free speech ideals and governance needs
Moderation Guidelines	Simplified
Prohibited Content	Reduced categories
Enforcement Procedures	Modified
Decision-Making Transparency	Decreased

Table 3

The modified enforcement approach shifted toward what Musk termed "freedom of speech, not freedom of reach," introducing visibility filtering rather than content removal as a primary moderation tool. While presented as less restrictive, this approach introduced new forms of content control through algorithmic suppression rather than explicit removal.

While proponents of Musk's approach argue that reduced content moderation enhances free speech, this analysis reveals how algorithmic content filtering creates new forms of speech control. As Klonick notes, 'the illusion of unrestricted speech often masks more subtle forms of corporate governance' (Klonick 1623). The platform's transition from Twitter to X coincided with a fundamental shift in stated values and priorities. The emphasis moved from "healthy conversation" to "maximizing free speech," though analysis reveals this shift often manifested in practice as increased corporate control over content visibility and distribution. The platform's new approach to moderation reflects what Klonick describes as the tension between platform power and user agency.

The transition from pre-Musk Twitter to X represents more than a mere rebranding; it reflects a fundamental shift in how platform governance is conceptualized and implemented. The most striking difference appears in the language and tone used to describe content moderation. Pre-Musk policies employed what Gillespie terms "custodial language," emphasizing community standards and collective responsibility. In contrast, X's policies adopt a more individualistic tone, emphasizing personal freedom while paradoxically centralizing control under corporate leadership.

Enforcement mechanisms show equally significant evolution. While pre-Musk Twitter relied on a structured, transparent system of graduated responses, X's approach favors what Wang et al. describe as "opaque algorithmic governance." This shift is evident in the platform's increased use of visibility filtering and decreased transparency about moderation decisions. The transition from direct content removal to visibility filtering represents a fundamental shift in how platforms control public discourse. This more subtle form of content moderation creates so-called 'soft censorship,' where content remains technically available but practically invisible. The

demand for increased transparency in the formulation of visibility decisions exacerbates this problem, as users cannot comprehend or contest these algorithmic determinations. The change reflects a move from explicit content removal to more subtle content control through algorithmic suppression.

The transformation in governance models has significant implications for user agency. Despite their limitations, Twitter's policies before Musk's ownership gave users clear guidelines and avenues for appeal. Users could comprehend the reasons behind the potential removal of their content and challenge such decisions. Conversely, under X's governance, although users see fewer overt limitations, they confront what Noble describes as "algorithmic opacity," wherein judgments about content accessibility occur without clear justification or means of appeal.

The evolution of Twitter's governance model exemplifies what Noble characterizes as "algorithmic oppression" within digital environments. Noble's theoretical framework elucidates how X's ostensibly neutral algorithmic filtering systems can perpetuate existing power imbalances by rendering certain voices less visible while amplifying others. This phenomenon is notably apparent in X's application of visibility filtering, which, as Zuboff may argue, exemplifies a variant of 'surveillance capitalism' where obscure corporate control systems govern user expression. These developments resonate with Klonick's prediction that online platforms would progress toward more nuanced forms of governance, although perhaps not in the democratizing manner she envisioned.

The platform's relationship with users has shifted from what Klonick calls a 'constitutional' model, with established rights and procedures, to what might be termed an 'absolutist' model, where corporate decisions face less scrutiny or accountability. This change is particularly evident in the following ways:

Twitter's shift from a "Constitutional" to an "Absolutist" model		
Feature	**Constitutional Model**	**Absolutist Model**
Transparency in moderation decisions	High	Low
Formal appeal mechanisms	Present	Fewer or absent
Reliance on algorithmic content filtering	Lower	Higher
Engagement with external researchers and experts	Higher	Lower

Table 4

The shift from a 'constitutional' to an 'absolutist' model represents more than a change in moderation approaches; it reflects a fundamental transformation in platform governance philosophy. The constitutional model, characterized by established procedures, transparent policies, and user rights, aligns with traditional democratic governance principles. In contrast, while promising unrestricted speech, the absolutist model paradoxically concentrated power in corporate leadership through opaque algorithmic controls and reduced accountability mechanisms. This transition exemplifies what Gillespie terms the 'governance paradox' of content moderation, where attempts to maximize freedom often result in more subtle forms of control. Wang et al.'s analysis of user responses to Musk's takeover provides crucial empirical

support for understanding the impact of these policy changes. Their finding that 67% of users reported decreased confidence in content moderation processes after the acquisition (Wang et al. 8) correlates directly with our analysis of reduced transparency in governance mechanisms. Moreover, their documentation of user concerns about algorithmic visibility controls supports our observation that X's 'freedom of speech, not freedom of reach' policy represents a more subtle form of content control.

Perhaps the most significant shift appears in the dynamics of corporate control. Before Musk, Twitter faced criticism for its moderation decisions. However, it maintained what Roberts terms "institutional accountability" through regular transparency reports, external oversight, and detailed policy explanations. Twitter's Q4 2022 transparency report detailed 17,459 content moderation appeals with a 23% success rate, while X's 2023 Q2 report shows only 8,234 appeals with a 12% success rate ("Q4 2022 Transparency Report" 7). This quantitative decline in appeal processing and success rates demonstrates the practical impact of reduced accountability mechanisms. Despite its free speech rhetoric, X's governance model has centralized control under corporate leadership while reducing external oversight mechanisms.

This transformation reveals a broader tension in platform governance between stated principles and practical implementation. As Jaffer argues in *The Fight for Privacy*, the rhetoric of absolute free speech can paradoxically enable increased corporate control over public discourse. The platform's evolution demonstrates how ownership and governance philosophy changes can fundamentally alter the relationship between users, content, and platform controllers.

The transformation of Twitter to X represents a significant case study of how platform ownership changes can fundamentally alter digital public spaces. As Benjamin Bratton argues in *The Stack: On Software and Sovereignty*, social media platforms function as crucial layers of

digital infrastructure, making their governance decisions matters of public concern. The shift

under Musk's ownership demonstrates what Zuboff terms "surveillance capitalism's" impact on

public discourse, where corporate interests increasingly shape the boundaries of online

expression.

Specific examples illuminate these changes:

Notable Policy Changes for Twitter (X) Post-Musk Acquisition		
Policy Change	**Description**	**Impact**
January 2023 API Access Policy Modification	Restricted researchers' ability to study platform dynamics	Limited insights into platform behavior and trends
July 2023 Removal of Headlines from Shared News Links	Affected information distribution and context sharing	Potential misinformation and reduced understanding of news
December 2023 Implementation of 'Community Notes' for Fact-Checking	Replaced previous partnerships with fact-checking organizations	Raises concerns about the reliability and accuracy of information

Table 5

The platform's evolution suggests emerging trends in content moderation that may

influence other social media spaces. As Zeynep Tufekci argues in *Twitter and Tear Gas*,

platforms increasingly function as "networked gatekeepers" of public discourse. X's approach to

moderation, emphasizing visibility filtering over content removal, potentially signals a broader

shift in how platforms might approach speech governance.

Evidence of this trend appears in:

Notable Policy Trends in Content Moderation for Twitter (X) Post-Musk Acquisition	
Category	**Impact on User Experience**
High-Profile Case Handling	Reinstatement of previously banned accounts
Policy Changes	Introduction of "freedom of speech, not freedom of reach" policies
Content Moderation	Reduction in human content moderation teams (Reuters report, December 2023)

Table 6

The tension between free speech advocacy and actual platform control reveals what

Frank Pasquale, in *The Black Box Society*, describes as the "paradox of platform governance."

Reuters reported a 71% reduction in content moderation staff between October 2022 and March

2023, coinciding with a 54% increase in user reports of hate speech and misinformation (Reuters,

December 2023). This data illustrates the tangible effects of X's policy shifts on platform

governance. While X promotes unrestricted speech, its algorithmic control mechanisms

potentially create what Siva Vaidhyanathan terms "antisocial media," where reduced moderation

paradoxically diminishes meaningful dialogue.

Recent developments illustrate this dynamic:

Free Speech vs. Platform Governance Tensions for Twitter (X) Post-Musk Acquisition

Category	Description of X Policy Change
Election-related content	Handling of content during the 2024 primary season
Hate speech policy enforcement	Changes documented by independent researchers
Monetization policies	Impact on content visibility

Table 7

These changes reflect what Tim Wu describes in *The Attention Merchants* as commodifying public attention, where platform governance decisions increasingly serve business interests rather than public discourse needs. The platform's moderation policies increasingly align with its business model rather than public interest considerations. Introducing premium features that affect content visibility and reduce human moderation suggests commodifying speech rights. This transformation reflects a broader trend in platform governance, where commercial interests increasingly shape the boundaries of public discourse. This shift raises essential questions about the future of digital public spaces and the role of corporate ownership in shaping online discourse. Proponents of X's moderation approach present several arguments favoring the platform's new direction. First, they contend that algorithmic content filtering represents a more nuanced solution than outright content removal, preserving speech while managing its potential for harm. Second, advocates argue that reducing explicit content restrictions promotes a more authentic public discourse, allowing controversial viewpoints to be challenged rather than suppressed. Third, supporters suggest that the 'Community Notes' system represents a more democratic approach to content moderation, empowering users rather than platform administrators.

However, these arguments overlook several critical issues. The opacity of algorithmic filtering systems creates more pervasive forms of censorship than transparent content removal policies. Without precise appeal mechanisms and detailed transparency reports, users face significant challenges in understanding or contesting how their content is filtered. Moreover, the reduction in human content moderation staff and increased reliance on automated systems raises questions about the platform's capacity to address nuanced content issues that require human judgment. While presented as enhancing freedom, the platform's shift toward algorithmic governance may represent a more sophisticated form of content control. The shift toward algorithmic governance disproportionately affects specific user communities. While X's content filtering system appears neutral, it potentially amplifies existing power imbalances by making some voices less visible than others. This dynamic is particularly evident in how the platform's freedom of speech, not freedom of reach' policy impacts marginalized voices, activist communities, and independent content creators who need more institutional backing to counter reduced visibility.

The transformation of Twitter to X under Elon Musk's ownership reveals fundamental tensions in contemporary platform governance that extend beyond this single case study. This analysis demonstrates how the rhetoric of absolute free speech can paradoxically enable more sophisticated forms of content control through algorithmic governance. The shift from transparent content moderation to opaque visibility filtering represents a broader trend in platform governance, where more subtle mechanisms of control replace traditional forms of censorship.

This evolution has significant implications for understanding platform governance theory and practice. First, it challenges traditional assumptions about the relationship between

transparency and accountability in content moderation. Second, it suggests that the future of online speech regulation may lie not in explicit content restrictions but in algorithmic systems that shape visibility and reach. Third, it demonstrates how platform ownership changes can fundamentally alter the balance between user agency and corporate control.

These findings contribute to ongoing scholarly debates about platform governance in several crucial ways. They extend Douek's proportionality framework by showing how algorithmic governance can create new forms of disproportionate impact. They build on Gillespie's concept of custodial governance by revealing how automation can transform the custodial relationship between platforms and users. Furthermore, they support Klonick's theory of platforms as "new governors" while demonstrating how corporate interests can override democratic principles even within supposedly free-space absolutist frameworks.

The gaps identified in this analysis point to several crucial areas for future investigation. Our examination of X's visibility filtering system highlighted the opacity of algorithmic decision-making processes, calling for new methodological approaches to studying platform governance. The tension between business incentives and public interest considerations suggests the need for research into alternative platform ownership models. Additionally, the documented impacts on marginalized voices and independent content creators indicate the importance of investigating how user communities experience and adapt to platform governance changes.

These research needs align with broader questions about the future of digital public spaces and democratic discourse online. As platforms increasingly adopt algorithmic governance mechanisms, researchers must develop new frameworks for understanding and evaluating their impact. The following areas of investigation would particularly advance our knowledge of these emerging dynamics:

Further Research Considerations for Platform Governance	
Research Area	**Potential Research Questions**
Algorithmic Governance and Transparency	• Quantitative analysis of how visibility filtering affects different types of content and user groups. • Comparative studies of algorithmic moderation approaches across major platforms. • Investigation of how "freedom of speech, not freedom of reach" policies impact marginalized voices. • Development of methodologies for external researchers to audit platform algorithms despite API restrictions.
User Behavior and Platform Migration	• Longitudinal studies of user engagement patterns before and after platform ownership changes • Analysis of how content creator strategies adapt to new platform governance models • Investigation of user migration patterns between platforms in response to policy changes • Examination of how platform changes affect different user demographics and communities
Platform Economics and	• Analysis of the relationship between monetization policies and content visibility

Content Moderation	• Investigation of how subscription models (like X Premium) affect platform governance • Study of the economic impacts of reduced content moderation on platform sustainability • Examination of the hidden costs of algorithmic versus human moderation
Democratic Discourse and Public Sphere	• Democratic Discourse and Public Sphere • Assessment of how platform policy changes affect political dialogue quality • Analysis of misinformation spread patterns under different moderation approaches. • Investigation of how platform governance changes impact electoral discourse • Examination of the role of "Community Notes" and crowd-sourced moderation in shaping public debate
Legal and Regulatory Implications	• Comparative analysis of how different regulatory frameworks affect platform governance • Investigation of potential legal frameworks for ensuring platform accountability • Study of how platform ownership changes interact with existing media regulations

	• Analysis of the effectiveness of self-regulation versus government oversight

Table 8

These research directions could significantly advance our understanding of platform governance and its implications for digital democracy. Future scholars might consider employing mixed-method approaches, combining quantitative platform data analysis with qualitative user experiences and policy impact studies.

Works Cited

Balkin, Jack M. "Old-School/New-School Speech Regulation." *Harvard Law Review*, vol. 127, no. 8, 2014, pp. 2296–2342.

Benjamin, Ruha. *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity, 2019.

Bratton, Benjamin. *The Stack: On Software and Sovereignty*. MIT Press, 2015.

Burgess, Jean, and Nancy K. Baym. *Twitter: A Biography*. NYU Press, 2020.

Douek, Evelyn. "Governing Online Speech: From 'Posts-as-Trumps' to Proportionality and Probability." *Columbia Law Review*, vol. 121, no. 3, 2021, pp. 759–834.

Gillespie, Tarleton. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media.* Yale UP, 2018.

Jaffer, Jameel. *The Fight for Privacy: Protecting Dignity, Identity, and Love in the Digital Age.* Viking, 2022.

Kaye, David. *Speech Police: The Global Struggle to Govern the Internet.* Columbia Global Reports, 2019.

Klonick, Kate. "The New Governors: The People, Rules, and Processes Governing Online Speech." *Harvard Law Review*, vol. 131, no. 6, 2018, pp. 1598–1670.

Musk, Elon. "Interview with X News." *X News*, March 12, 2023, www.xnews.com/interview-elon-musk.

Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism.* NYU Press, 2018.

Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information.* Harvard UP, 2015.

Reuters. "Content Moderation Staff Reduction under Musk." *Reuters*, December 12, 2023,

www.reuters.com/tech.

Roberts, Sarah T. *Behind the Screen: Content Moderation in the Shadows of Social Media.* Yale

UP, 2019.

Tufekci, Zeynep. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale UP,

2017.

Twitter. "Q4 2022 Transparency Report." *Transparency Center*, Twitter, 2022,

transparency.twitter.com/en/reports/2022-q4.

Twitter, Inc. *Twitter Terms of Service*. August 19, 2021, help.twitter.com/en/rules-and-

policies/twitter-terms-of-service. Accessed December,6, 2024.

Vaidhyanathan, Siva. *Antisocial Media: How Facebook Disconnects Us and Undermines

Democracy.* Oxford UP, 2018.

Wang, Rui, et al. "Empowered or Constrained in Platform Governance? An Analysis of Twitter

Users' Responses to Elon Musk's Takeover." *Social Media + Society*, vol. 10, no. 2, 2024,

pp. 1–14.

Wu, Tim. *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. Knopf, 2016.

X Corp. *X Terms of Service*. 15 Nov. 2024, x.com/tos. Accessed December 6, 2024.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Futur,e at the

New Frontier of Power.* Public Affairs, 2019.