The Contagion of Distant Politics: Partisan Discourse on Chinese-language Social Media During Major U.S. Political Events

Muzhi Liu, Columbia University

June 21, 2025

Abstract

This study examines how major partisan events in the United States shape information ecosystems on Chinese-language social media platforms. Analyzing 152,606 LLM-labeled posts from X and 33,108 from Weibo during the July 2024 Trump assassination attempt, alongside over 9,200 posts and 200,000 comments from RedNote during the November 2024 presidential election, I document several key findings. First, major political events trigger asymmetric content creation: influencers dramatically increase posting while general users show modest increases, with all groups producing more misinformation and partisan content during these events. Second, these events and their associated information flows shape downstream online behavior of Chinese-speaking social media users. Specifically, these linguistically remote users exhibit partisan behavior that mirrors patterns observed in U.S. contexts: Users engage more frequently with partisan content while spreading partisan messages and misinformation themselves. Their participation also shows patterns of affective polarization and selective exposure. These effects transcend geographic boundaries but attenuate after the resolution of the events. The findings demonstrate that U.S. political polarization operates as a transnational phenomenon, transforming information quality and discourse norms among linguistically distant observers through platform-mediated mechanisms.

Keywords: partisan identity, misinformation, social media, Chinese-language users, U.S. elections

Word Count: 8,257

Acknowledgments: I thank Shigeo Hirano, Donald Green, Eunji Kim, Yamil Velez, Carlo Prato, Naoki Egami, Junyan Jiang, Patrick Liu, Alec Ewig, Ronald Wu, Hanying Wei, Yinghui Zhou, Lesley Zhang, and Emma Swanson for valuable comments and suggestions, as well as participants in the Comparative Politics Seminar at Columbia University. This project was recognized as the first-place winner of the 2025 Young Public Opinion Scholar Competition by the New York and Pennsylvania-New Jersey chapters of AAPOR. Any errors remain my own.

Author Information: Muzhi Liu is a PhD student in the Department of Political Science at Columbia University. Email: ml4967@columbia.edu

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Introduction

On July 13, 2024, gunshots rang out at a Donald Trump rally in Pennsylvania, marking the first assassination attempt on a U.S. presidential candidate in decades. Within hours, Chinese-language social media erupted with speculation, conspiracy theories, and partisan interpretations of the event. Users who had never set foot in the United States debated whether the attack was staged, analyzed ballistic trajectories, and shared memes portraying Trump as divinely protected. This explosion of engagement among linguistically and geographically distant observers raises an interesting puzzle: how do political shocks in one country shape information ecosystems among another group of people, who speak a different language, perhaps come from another country, and are perhaps merely "electoral spectators" who may not have a direct stake in election?

This paper examines the transnational influence of highly partisan political discourse by analyzing how Chinese-speaking social media users process and propagate information about U.S. electoral events. While extensive research documents how misinformation spreads within national borders (Grinberg et al., 2019; Vosoughi et al., 2018) and online identity-charged partisan behavior (Hopkins et al., 2024), we know remarkably little about cross-linguistic transmission of political content. This gap matters because digital platforms create unprecedented opportunities for political events to influence global audiences who lack direct stakes in electoral outcomes. For audiences with limited prior experience with elections, exposure to U.S. political discourse becomes an opportunity to understand an important component of world politics. Therefore, characterizing how these events shape their political understanding is vital.

I analyze Chinese-language discourse surrounding two pivotal moments in the 2024 U.S. presidential cycle: the July assassination attempt on Donald Trump and the November presidential election. This empirical strategy leverages exogenous shocks to identify causal effects of political events on information quality and user behavior. Chinese-speaking users provide an ideal population for studying "distant witnesses"—individuals who observe and discuss foreign politics without electoral participation rights. Their engagement patterns reveal how political polarization operates as a communicable phenomenon, transmitted through digital networks across linguistic and cultural boundaries.

Using Large Language Models (LLMs) to analyze around 150,000 posts from X, 33,000 posts from Weibo, as well as over 9,000 posts and 200,000 comments from Rednote, I document several key mechanisms of transnational political influence. First, political events trigger asymmetric content creation, with influencers dramatically increasing posting while general users show more modest increases, with all groups producing more misinformation and partisan content during these events. Second, these events and their associated information flows shape downstream online behavior of Chinese speakers, leading them to engage more frequently with partisan content following these events. Third,

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

consequently, these linguistically remote users exhibit online behavior that mirrors partisan behavior in the United States, spreading partisan messages and misinformation themselves. Their engagement also exhibits patterns of affective polarization and selective exposure. These effects transcend geographic boundaries but attenuate after the resolution of the events. These findings extend our understanding of political communication in three ways. Theoretically, they demonstrate that political polarization may operate through universal psychological mechanisms that transcend cultural contexts. Methodologically, the paper introduces a cross-platform identification strategy that exploits variation in content moderation policies to isolate exposure effects. Substantively, the results reveal how U.S. domestic politics shapes global information environments, with implications for democratic discourse worldwide.

The paper proceeds as follows. Section 2 reviews literature on online spread of partisan information and misinformation, and cross-cultural political communication. Section 3 describes the data collection and computational methods. Sections 4 present results from the assassination attempt and election studies respectively. Section 5 presents the conclusion and discusses the implications from this study's findings.

Literature and Theory

Understanding how Chinese-speaking social media users engage with U.S. political events requires synthesizing insights from several distinct yet interconnected bodies of literature: (1) online political communication and partisan identity formation, (2) information dynamics during major political events, and (3) cross-linguistic and transnational information flows. While each domain offers valuable perspectives, a critical gap exists in understanding how these phenomena intersect when linguistically and culturally distant audiences engage with foreign political shocks. This section outlines the established findings in these areas and develops theoretical expectations for how U.S. political events might cascade through Chinese-language information ecosystems.

Partisan Identity and Online Behavior

Political content creation on social media is driven by identity-centered motivations, as users preferentially respond to and spread information that resonates with their core identities (Hopkins et al., 2024). In U.S. contexts, partisanship functions as a powerful social identity where political affiliation becomes a lens for group belonging and out-group stereotyping (Green et al., 2002). This identity-based partisanship fosters in-group cohesion while generating out-group animosity (Iyengar et al., 2019; Iyengar and Westwood, 2015; Mason, 2015), leading to more emotional and intense political engagement (Huddy et al., 2015).

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

These psychological processes manifest in predictable online behaviors through selective exposure mechanisms. Users engage in systematic information segregation, selectively consuming content that aligns with their pre-existing beliefs (Gentzkow and Shapiro, 2011) while actively curating digital environments to minimize exposure to challenging viewpoints (Green et al., 2025). This behavior extends across multiple dimensions of online engagement: users preferentially follow like-minded influencers (Mukerjee et al., 2022), engage primarily with ideologically similar peers (Barberá et al., 2015), and interact with sources that reinforce their worldviews (Nyhan et al., 2023). On the other hand, when inadvertently exposed to opposing perspectives, users may experience backfire effects that paradoxically strengthen rather than moderate their original beliefs and may heighten their hostility toward perceived out-groups (Bail et al., 2018). This out-group hostility can translate to "more extreme" online behavior: individuals holding strong negative feelings toward political opponents are more likely to use toxic language (Falkenberg et al., 2024) and share content that involves misinformation to attack their political out-groups (Osmundsen et al., 2021).

One theoretical foundation for these patterns lies in motivated reasoning, which distinguishes itself from Bayesian updating through information processing biases. While Bayesian updaters incorporate new information regardless of its inclinations in ways that follow Bayes' rule, motivated reasoners selectively attend to and process information that confirms their pre-existing perspectives (Little, 2025). When confronted with counter-attitudinal information, motivated reasoners either fail to update their beliefs or, paradoxically, update in the opposite direction (Bartels, 2002). However, recent experimental evidence suggests that such backlash effects occur less frequently than earlier research indicated, typically emerging only under specific conditions—particularly when content is highly provocative and directly challenges users' core political identities (Guess and Coppock, 2018; Velez and Liu, 2024).

Scholars have noted that these patterns may not extend universally across political contexts. Kobayashi et al. (2024) examine partisanship in Japan and Hong Kong, finding that users exposed to partisan messages do not exhibit the same selective exposure patterns observed in American contexts. The authors attribute this absence to weaker affective polarization, especially an absence of outgroup hostility. Importantly, Kobayashi et al. (2024) use information related to the countries' own party systems as treatments in their survey experiments, like the party systems in Hong Kong and Japan, while these countries' party systems may not be as encompassing as a mega-identity as in American politics (Mason, 2015). Their findings, while effectively answering how domestic partisanship operates as an identity in non-U.S. contexts, introduce an interesting but unanswered question: how U.S. partisan mechanisms—a system that is highly encapsulative of identities across many dimensions—operate when partisan cues flow overseas and are interpreted by linguistically and culturally distant audiences.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

"Distant Witnesses" During Major Political Events

Major political events function as natural experiments for understanding how information ecosystems respond to exogenous shocks. The focusing events literature demonstrates that sudden, unexpected, and highly visible events can elevate event-relevant issues to the top of the public agenda and catalyze shifts in public opinion (Birkland, 1998).

Two recent studies illustrate this pattern. Research on the 2020 George Floyd protests found that this exogenous event rapidly decreased police favorability and increased perceptions of anti-Black discrimination among low-prejudice and politically liberal Americans, while attitudes among high-prejudice and politically conservative Americans remained largely unchanged or showed only small, ephemeral shifts (Reny and Newman, 2021). Similarly, analysis of the July 2024 Trump assassination attempt revealed that Republicans, including MAGA supporters, became significantly less supportive of partisan violence against Democrats following the event, and they did not develop greater hostility toward Democratic partisans (Holliday et al., 2024). These findings, while confirming that learned political attitudes are deeply ingrained and generally resistant to persuasion (Krosnick and Petty, 1995; Sears, 1993), nevertheless show that exogenous events can produce detectable and durable effects even among those who are not so predisposed to update their belief in the direction suggested by the event.

Political information increasingly traverses national and linguistic borders via digital platforms, yet research on transnational information flows has predominantly focused on elite actors (Elkins and Simmons, 2005) or state-sponsored campaigns (King et al., 2017) rather than organic citizen engagement. The phenomenon of ordinary citizens engaging with foreign political events as "distant witnesses" (Chouliaraki, 2016) remains underexplored, particularly regarding non-Western audiences observing political developments in powerful nations like the United States.

Political shocks present a unique opportunity to examine this dynamic more systematically. When major political events occur, information flows rapidly across borders to geographically and linguistically remote populations, transforming distant audiences into active spectators who not only consume but also engage with foreign political content. This transnational engagement raises important questions about how political information flowing from its country of origin may shape opinions and behaviors among remote populations who lack direct stakes in the outcomes but nonetheless participate in global political discourse. Recent studies reveal significant vulnerabilities for non-English speakers online. Spanish-speaking communities face disproportionate targeting by misinformation, with fact-checking mechanisms performing less effectively in Spanish (Abrajano et al., 2024; Velez et al., 2023). This linguistic vulnerability reflects broader systemic issues where platforms invest fewer resources in non-English content moderation (Pérez and Tavits, 2022). These findings highlight how language barriers can both

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

exclude communities from corrective information flows and make them more susceptible to targeted misinformation.

My study addresses a key gap in this literature by examining how partisan information affects a linguistically remote audience—Chinese speakers engaging with U.S. political content. While existing research has explored the relationship between partisanship and online behavior, and how linguistically distant groups respond to misinformation, no previous study has investigated how exposure to partisan information shapes the online discourse of "distant witnesses."

Theory and Contributions

The Chinese-speaking online sphere presents a particularly complex case due to the interplay of state information control and diverse diaspora experiences. Within mainland China, the information ecosystem is shaped by state censorship and curated narrative promotion (King et al., 2013, 2017; Lu et al., 2025; Xu, 2021; Xu et al., 2022). The state systematically monitors and removes sensitive content while disseminating "cheerleading messages" to align public sentiment with state interests, potentially suppressing organic political expression through "strategic distraction" (King et al., 2017; Lu et al., 2025). Outside mainland China, the global Chinese diaspora utilizes diverse platforms—from global ones like X to Chinese-origin platforms like Weibo and Rednote—with information consumption patterns influenced by both host country media environments and ties to Chinese-language media (Sun and Yu, 2023).

This constrained information environment creates a unique dynamic: major foreign political events often serve as rare opportunities for politically relevant discussion on Chinese-speaking social media platforms. Since such discussions are typically suppressed or absent, foreign political shocks may generate unusually strong agenda-setting effects among Chinese-speaking users who have latent interests in political engagement. This leads to my first set of hypotheses:

> **H1:** Major U.S. political events will have agenda-setting effects among Chinese-speaking users, with the magnitude of these effects increasing with users' exposure to more event-related information online.
>
> **H2:** Agenda-setting effects from U.S. political events will increase Chinese-speaking users' political engagement, manifesting as greater attention to political discussions, increased participation in online political discourse, and heightened willingness to express political attitudes publicly.

The cases examined—the Trump assassination attempt and the 2024 U.S. presidential election—are both highly partisan events. Following agenda-setting theory, exposure

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

to such polarized information should make "partisan identity" more salient for Chinese-speaking users, even when such events would normally be outside their direct concern. While many Chinese-speaking users lack formal partisan affiliations or voting rights in U.S. elections, their pre-existing worldviews can still be mapped onto the American political spectrum. Therefore, they selectively process information that reinforces their beliefs and actively facilitate the spread of such information (Zaller, 1992). This reasoning generates my third hypothesis:

> **H3:** Chinese-speaking users will identify with partisan camps and engage in behaviors characteristic of U.S. partisanship: partisan cheerleading, affective polarization, misinformation spread, attack language usage, and echo chamber formation.

Beyond these theoretical contributions, this study makes several important empirical and methodological advances:

- **Expanding geographic scope:** This research focuses on Chinese-speaking social media users as a globally significant yet often overlooked non-Anglophone community engaging with U.S. politics, moving beyond the typical focus on Spanish-language communities in U.S.-centric research.
- **Leveraging major events:** By examining reactions to major U.S. political events, this study captures how foreign political shocks trigger changes in information creation and consumption patterns among distant audiences.
- **Cross-platform comparative analysis:** The study investigates how varying platform architectures, user demographics, and regulatory environments (across X, Weibo, and Rednote) shape the transnational flow and reception of political information and misinformation. Such comparative approaches remain rare in studies of distant witnessing.

Data and Methods

Data Collection

This paper examines X, Weibo, and Rednote as primary platforms of interest. These platforms facilitate political discussion through their architectural design and low-tie environments (Theocharis et al., 2022). Specifically, X, Weibo, and Rednote enable users to form and maintain networks based on shared interests or weak ties without requiring mutual consent to establish connections (Bossetta, 2018). In contrast, platforms like Facebook and WeChat are designed to facilitate communication within networks where

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

strong ties are relatively more important and relationships are often grounded in offline connections (Vaccari and Valeriani, 2021).

X and Weibo share remarkably similar interfaces and functionalities: both platforms allow users to post short messages with character limits and restricted image uploads, while enabling other users to like, comment, and repost content. Rednote differs in key ways—it permits longer posts (up to 1,000 words) and more images (up to 18 per post) but prohibits reposting, limiting user interactions to likes and comments only.

These platform differences shape my analytical approach. I compare Weibo and X in Study 1 (July 2024) because their "tweet-driven" architecture—where comments can spawn new posts—facilitates the observation of downstream behavioral spillovers. This design makes these platforms ideal for testing my hypotheses about partisan attitude formation and information propagation, including misinformation spread, affective polarization, and attack language usage. In contrast, I use Rednote data for Study 2 (November 2024) to leverage its distinct interaction structure. Since Rednote features tighter coupling between original posts and comments, and users cannot repost content, commenting behavior more directly reflects selective exposure patterns. This environment allows me to examine the echo chamber part of my hypothesis, as users' commenting choices reveal their preference for engaging with ideologically aligned content.

The user demographics of these three platforms differ significantly. Weibo users are predominantly from mainland China. While Rednote's user base is also primarily mainland Chinese, it hosts a substantial international Chinese-speaking population. X has the smallest proportion of mainland Chinese users, as direct access to the platform is blocked in China, requiring VPN usage. These demographic differences necessitate caution when comparing user behavior across platforms. Two primary concerns arise: First, do users on international platforms have a higher baseline political knowledge, interests, or engagement? This selection problem may be particularly pronounced when we compare Weibo to X in study 1, because users who self-select to X may be intrinsically more politically sophisticated or engaged. Second, given the different demographic compositions of Chinese-speaking users, how do we know which proportions of users are "driving up" the conversation?

I address these methodological challenges through selective data collection approaches. To account for X's larger international user base, I exclusively collected Chinese-language posts and comments, treating "posting in Chinese" as a unifying characteristic across platforms. While most Weibo and Rednote users post in Chinese by default, Chinese-language users represent a smaller subset on X, potentially making the populations more comparable for Study 1. Although this approach cannot fully eliminate selection bias—Chinese-language users on X may still differ from those on other platforms in baseline political attitudes and engagement levels—verification of parallel trends in subsequent sections helps alleviate this concern.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

I employed keyword-based data collection for both studies. In Study 1, I collected posts from Weibo and X using a shared keyword list, spanning 10 days before and 14 days after the July 2024 assassination attempt on Trump. I collected posts from X using X's API.[1] For Study 2, I collected Rednote posts and comments using targeted keywords, spanning two weeks before and after the November 2024 U.S. presidential election.[2] A full list of keywords is shown in Table 1 below. After collecting data, I filtered out posts and comments that contained the target keywords but were unrelated to politics. The removed content from X primarily discussed blockchain markets, while the removed content from Weibo focused on Chinese celebrity and pop culture topics, including TV episodes, shows, and fictional novels.

Table 1: Keywords Used to Collect Comments

Study	Chinese Keywords	English Translations
Study 1	川普, 特朗普, 暗杀, 枪手, 投票, 共和党	Trump (two different expressions), assassination, gunman, vote, the Republican Party
Study 2	美国大选, 川普, 特朗普, 哈里斯, 拜登, 民主党, 共和党, 投票	US Election, Trump (two expressions), Harris, Biden, Democrat, Republican, vote

To address the second concern, I included IP locations of users on each platform. This enables me to conduct subgroup analyses within users from a specific geographical area - like mainland China, the U.S., or other countries.

Constructing Variables through LLM

One important component of this study is conducting content analysis of scraped content through Large Language Models (LLM).

Figure 1: Procedures and Examples of Text Processing

Following text collection, I implement standard preprocessing procedures, including URL removal and common emoji filtering. I then deploy a LLM for two primary content

[1] I registered for Pro Service of X's API for a month, as only this user tier can access archived posts and allows for 1 million post retrievals per month. See example code for utilizing X's API to collect posts in Appendix A.3.

[2] Given Rednote's platform architecture, data collection was performed through continuous monitoring of recent posts rather than archived retrieval, capturing content within seven days of publication to maintain temporal relevance to the election timeline.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

analysis tasks: (1) automated labeling of posts and comments, and (2) standardization of user IP locations. To protect user privacy, I anonymized all posts and comments by assigning unique identifiers to each user before analysis. The LLM processed only three data fields: post or comment content, users' IP locations, and user biographies.

The content labeling system applies multiple categorical labels organized into three conceptual domains. The information type and quality domain captures content accuracy through an ordinal misinformation score (0 to 1 in 0.25 increments) and partisan orientation via binary indicators for pro-Republican, anti-Republican, pro-Democrat, and anti-Democrat sentiment. The partisan identity and behavior domain measures candidate-specific support for each party's presidential nominee, affective polarization indicators (ingroup affinity and outgroup hostility), and the presence of attack language. The other attributes domain includes standardized geographic location data derived from user IP information.

IP location standardization applies exclusively to X data. Users on X self-report their locations, resulting in a mixture of genuine geographic information (e.g., "Pomona, California") and fabricated entries (e.g., "Dorm Room"). Using LLMs, I categorize these locations into four groups: China, United States, International (users from countries other than China or the U.S. with verifiable locations), and Undisclosed (users who either provide no location or submit fabricated information). Table 2 shows labeling schemes in detail.

It is important to note that misinformation, partisan inclinations, and partisan identities are often hard to label. Prior research in misinformation detection has relied primarily on binary categorizations or expert human labeling. Previously used approaches labeled content as originating from "untrustworthy websites" (Nyhan, 2020) or identified misinformation through fact-checker debunking (Allen et al., 2024; Mattozzi et al., 2022; Osmundsen et al., 2021). On the other hand, recent studies have demonstrated that fine-tuned pre-trained language models and LLMs like GPT-3.5 can achieve commendable results in fake news detection without extensive auxiliary data, with performance comparisons indicating that LLM-based detection systems achieve good accuracy compared to traditional rule-based approaches.

The theoretical foundations for LLM-based misinformation labeling rest on three key capabilities that address the limitations of previous approaches (Chen and Shu, 2024). First, LLMs contain significant amounts of world knowledge since they are typically pre-trained on large corpora such as Wikipedia and have billions of parameters, enabling them to store much more knowledge than single knowledge graphs and detect factual errors in misleading texts (Sun et al., 2023). Second, LLMs demonstrate strong reasoning abilities, especially in zero-shot contexts, with powerful capacities in arithmetic reasoning, commonsense reasoning, and symbolic reasoning (Huang and Chang, 2022), allowing them to decompose problems and reason based on rationales (Kojima et al., 2022). Third, LLMs

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Table 2: LLM Content Coding Framework

Variable	Coding Rules
Information Type and Quality	
misinfo	**0**: Accurate, well-evidenced, or widely accepted as true **0.25**: Generally accurate but omits crucial details or includes minor inaccuracies **0.5**: Partially misleading or lacking key context; some unverified claims **0.75**: Mostly inaccurate or strongly misleading, but not entirely absurd **1**: Blatantly false, conspiratorial, or extremely misleading with no factual basis
pro_dem	**1** if content praises, endorses, or defends the US Democratic Party; **0** otherwise
anti_dem	**1** if content criticizes or attacks the US Democratic Party; **0** otherwise
pro_rep	**1** if content praises, endorses, or defends the US Republican Party; **0** otherwise
anti_rep	**1** if content criticizes or attacks the US Republican Party; **0** otherwise
Partisan Identity and Behavior	
trump_support	**1** if comment explicitly expresses support for Trump; **0** otherwise
harris_support	**1** if comment explicitly expresses support for Harris; **0** otherwise
ingroup_like	**1** if comment mentions who the commenter's in-group is and displays an affinity toward the group; **0** otherwise
outgroup_hate	**1** if comment mentions who the commenter's out-group is and displays negative attitudes toward the group; **0** otherwise
attack	**1** if content includes direct insults or derogatory language toward individuals or groups; **0** otherwise
Other Attributes	
ip_coded	**China**: User's self-reported IP location is in China **US**: User's self-reported IP location is in the US **International**: User's self-reported IP location is real, but not in China or US **NA**: User's self-reported IP location is either empty or fabricated

Notes: All variables coded using large language model (LLM) automated content analysis. Examples in original language (Chinese) provided in coding instructions. Binary variables coded as 1 (present) or 0 (absent). Misinformation variable uses ordinal scale from 0 (accurate) to 1 (completely false).

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

can be augmented with external knowledge, tools, and multimodal information, with recent research showing that LLM hallucinations can be mitigated through augmentation with retrieved external knowledge (Li et al., 2022) or real-time search engines (Gou et al., 2023). Empirical studies across multiple domains including politics, public health, and finance have demonstrated that LLM-based approaches for identifying misinformation show superior effectiveness compared to text-based models that rely solely on linguistic signals (Huang et al., 2024).

My labeling scheme of misinformation employs a graduated scale adapted from PolitiFact's Truth-O-Meter system, which provides a more nuanced approach to misinformation labeling than binary classifications. PolitiFact's Truth-O-Meter uses six ratings in decreasing levels of truthfulness: True (accurate with nothing significant missing), Mostly True (accurate but needs clarification), Half True (partially accurate but lacks important details), Mostly False (contains truth but ignores critical facts), False (not accurate), and Pants on Fire (not accurate and makes ridiculous claims) (Adair, 2018). My five-point scale (0 = accurate and well-evidenced, 0.25 = generally accurate with minor issues, 0.5 = partially misleading, 0.75 = mostly inaccurate, 1.0 = blatantly false) maintains this graduated approach while providing the numerical precision necessary for quantitative analysis. My operationalization of affective polarization builds on Iyengar and Westwood (2015)'s conceptualization as "the tendency of people identifying as Republicans or Democrats to view opposing partisans negatively and copartisans positively." Recognizing that Chinese speakers may not maintain strict partisan identities comparable to American political affiliations, I adapt this framework to capture broader patterns of political group-based affect. Specifically, I code binary indicators for ingroup affinity and outgroup hostility, where a value of 1 is assigned if and only if: (1) the text explicitly identifies a relevant political outgroup (for outgroup hostility) or ingroup (for ingroup affinity), and (2) the text demonstrates clear negative attitudes toward the outgroup or positive attitudes toward the ingroup, respectively.

For study 1, I employ the newest cost-efficient model ChatGPT-4.1. For study 2, I employ ChatGPT-4o-mini due to its lower cost with massive text labeling. Through iterative prompt engineering and few-shot learning optimization, this model achieves both cost-effectiveness and reliable performance in text classification tasks, while ensuring that the usage is not heavily constrained by the API rate limit.[3] To validate the automated labeling process, I conduct expert annotation on a random sample of 500 comments from each dataset. This validation protocol serves dual analytical purposes. First, it enables systematic assessment of LLM labeling accuracy through correlation analysis between automated and expert annotations.[4] Second, it provides the expert-labeled ground truth necessary for implementing Egami et al. (2023)'s Design-Based Supervised

[3]Complete Python implementation and few-shot examples are provided in Appendix A.
[4]See this in Appendix B.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Learning (DSL) framework as a robustness check. As Egami et al. (2023) demonstrate, LLM-based labeling typically introduces systematic rather than random measurement error in constructed variables. Their DSL methodology, which leverages expert annotations to identify and correct these systematic biases, substantially improves the reliability of downstream statistical inference when using LLM-generated labels.

Table 3 presents descriptive statistics of the labeled datasets. On X, misinformation prevalence reaches 15.6%, while Weibo shows lower rates at 8.4%. Both platforms demonstrate substantial pro-Republican sentiment, with X showing 14.5% pro-Republican content compared to Weibo's 5.7%. Pro-Democrat content remains minimal across both platforms (1.8% on X, 1.1% on Weibo). This asymmetry suggests that the July 2024 assassination attempt triggered widespread partisan speculation and dissemination of unverified information across Chinese-language social media.

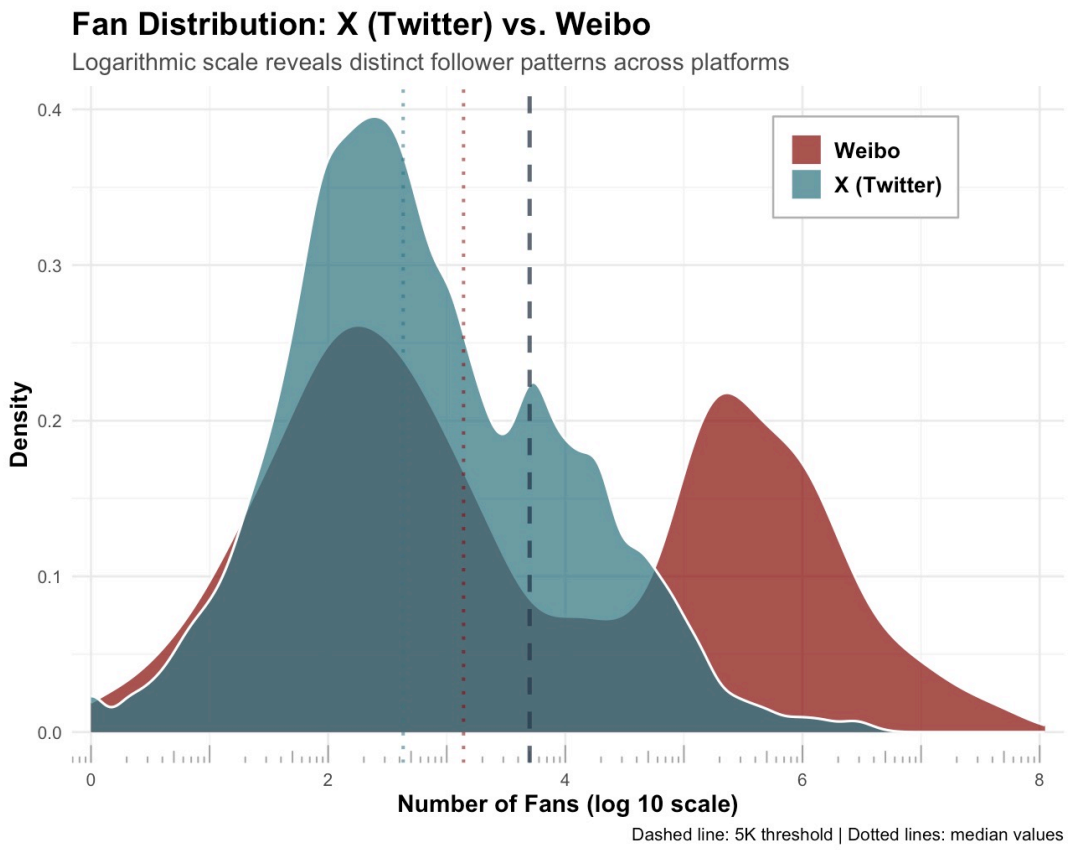


Figure 2: Follower Distribution of Users in Study 1

"Content creator" in Study 1 are defined as users with over 5,000 followers. Figure 2 demonstrates the rationale for this threshold. When follower counts are plotted on a log 10 scale, the distribution exhibits a bimodal pattern across both platforms. The first mode centers around 10^2 (100 followers) for both Weibo and X, while the second mode peaks at 10^4 (10,000 followers) for X and $10^{5.5}$ (316,000 followers) for Weibo. The 5,000-follower threshold strategically captures users in the second distribution mode—the "high-

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Table 3: Descriptive Statistics Across Platforms and Studies

Variable	Study 1		Study 2
	X	**Weibo**	**RedNote**
Content Analysis			
Misinformation prevalence	0.156	0.0842	0.154
	(0.270)	(0.208)	(0.260)
Pro-Republican attitude	0.145	0.057	0.077
	(0.352)	(0.232)	(0.267)
Pro-Democrat attitude	0.018	0.0113	0.027
	(0.133)	(0.106)	(0.161)
Attack messages	0.205	0.0837	0.091
	(0.404)	(0.277)	(0.287)
Total posts/comments	152,606	33,108	202,554
Content Creators			
Total unique creators	3,593	3,909	7,027
IP Locations			
China	408	3,112	3,601
United States	420	73	849
International	675	211	477
Undisclosed	2,090	513	2,100
Aggregate Profiles			
Political Profiles	5,445	199	–
Fans (mean)	94,628	2,259,638	4,184
Following (mean)	2,054	1,701	235.1
Commentators			
Total unique users	31,300	14,544	186,314
IP Locations			
China	2,191	9,002	104,042
United States	2,726	140	34,781
International	4,846	896	40,731
Undisclosed	21,537	4,533	6,760

Note: Standard deviations in parentheses. RedNote was formerly known as Xiaohongshu. Political profile indicates users with explicit political inclinations in their profile descriptions. Political profiles of note producers on Rednote were not collected.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

influence" segment—who likely exhibit distinct behavioral patterns due to their elevated reach and engagement potential. In Study 2, "Content creator" encompass all users who publish posts, reflecting the platform-specific characteristics of Rednote. Unlike the brief, comment-length posts typical of X and Weibo, content creation on Rednote requires substantially greater effort and investment, making any posting activity indicative of meaningful content creation behavior.

Study 2 data shows that RedNote messages exhibit a misinformation prevalence at 15.4%. While pro-Republican messages (7.7%) remain more prevalent than pro-Democrat messages (2.7%) on RedNote, the overall partisan content represents a smaller proportion compared to X's levels in Study 1. Cross-platform comparisons reveal significant variation in discourse civility. X demonstrates the highest prevalence of attack language (22.2%), substantially exceeding both Weibo (8.4%) and RedNote (9.1%). This disparity suggests that platform-specific factors—potentially including user demographics, moderation policies, or cultural norms—shape the tenor of political discussion among Chinese-speaking users.

Table 3 also presents geographical compositions of netizens on different platforms. X has way fewer mainland Chinese users, compared to Weibo and Rednote. Rednote hosts a much larger population of US and international Chinese-speaking netizens than Weibo.

Results

Study 1

In Study 1, I examine content creation patterns among influencers and general users during a 24-day window surrounding the July 13, 2024 assassination attempt on Donald Trump. The analysis spans 10 days before and 14 days after the event.

To examine differential content creation patterns following the assassination attempt, I aggregate message counts by unique username and categorize them as pre-event or post-event based on timestamps. I conduct subgroup analyses using coded IP locations, stratifying users into four categories: mainland China, United States, international, and undisclosed locations. To facilitate interpretation, I collapse the ordinal misinformation scale (0–1) into three categories: mostly accurate (scores $\in \{0, 0.25\}$), partially misleading (score $= 0.5$), and mostly inaccurate (scores $\in \{0.75, 1\}$). [5]

Treating the assassination attempt as an exogenous shock enables a quasi-experimental design. I estimate the following regression model to identify the causal effect of this salient political event on content creation behavior across user types and content characteristics:

$$\text{Comment Count}_{i,t} = \beta_0 + \text{Time}_t \times \text{Netizen}_i \times \text{Content}_i \times \text{IP}_i + \epsilon_{i,t} \quad (1)$$

[5] A time-series plot of content creation is in Appendix C.1.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

The model specification represents a full factorial design examining how content production varies across time periods, user types, content categories, and geographic locations. Time_t is a binary indicator distinguishing the post-event period from the baseline, capturing temporal shifts in user behavior following significant events. Netizen_i is a dummy variable identifying users with 5,000 or fewer followers (netizens) relative to the reference category of influencers with more than 5,000 followers. Content_i represents ordinal content categories that differ by platform and model: for misinformation models, this captures varying levels of misinformation content (Low, Medium, High) relative to the reference category of Mostly Accurate content, while for partisanship models, it distinguishes Left-leaning and Right-leaning partisan content from Neutral content. IP_i denotes user geographic location categories, which vary by platform due to different user bases and regulatory environments: X/Twitter models include China, International, and NA/Undisclosed locations relative to US users, while Weibo models compare International and Undisclosed locations to the reference category of China domestic users. The specification includes all possible two-way, three-way, and four-way interactions between these factors, allowing for complex conditional relationships in how different user types respond to events across content categories and geographic contexts. The error term $\epsilon_{i,t}$ incorporates robust standard errors clustered at the username level to account for within-user correlation in posting behavior over time.

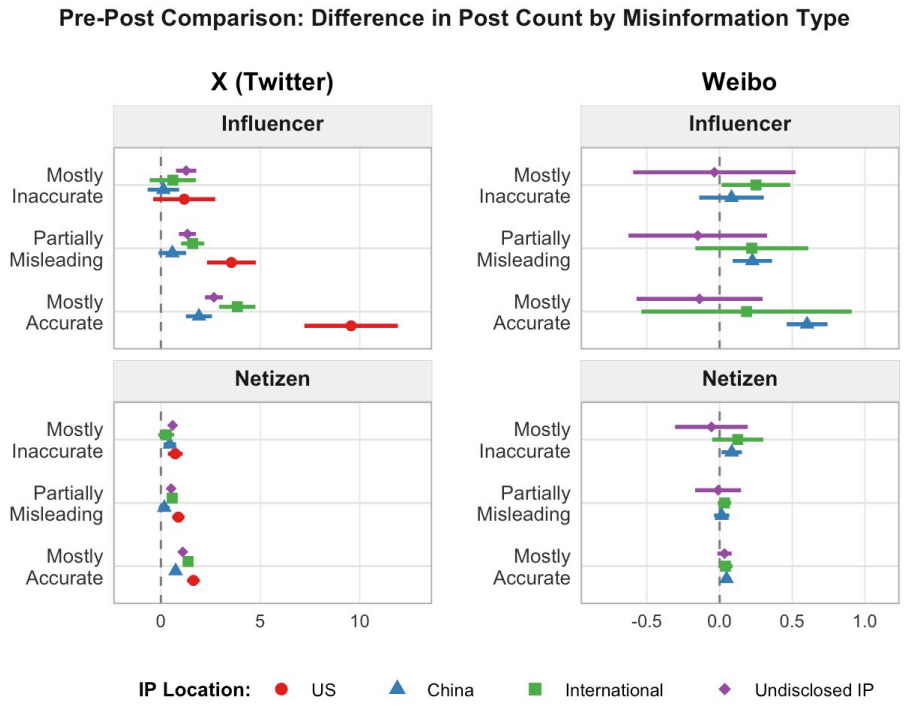
Figure 3 presents the estimated effects of the assassination attempt on Donald Trump on content creation patterns among Chinese-speaking users, stratified by user type (influencer versus general user) and IP location.[6] Panel A displays misinformation content creation, while Panel B illustrates partisan content patterns.

Across both platforms, influencers drive the post-event surge in content creation, consistently outpacing general users. On X, U.S.-based Chinese-speaking influencers demonstrate the most pronounced response. Following the assassination attempt, these influencers produce more mostly accurate messages than misinformation-containing content, with U.S.-based influencers exhibiting the highest activity levels. Compared to the pre-event period, U.S.-based influencers create approximately 10 additional mostly accurate messages and around 5 additional messages containing misinformation. Their partisan content production similarly intensifies: U.S.-based influencers generate approximately 8 more non-partisan messages, 4 more pro-Republican messages, and 3 more pro-Democratic messages on average during the post-event period. While influencers from other geographic locations demonstrate significant post-event increases, their response magnitude remains substantially smaller than their U.S.-based counterparts.

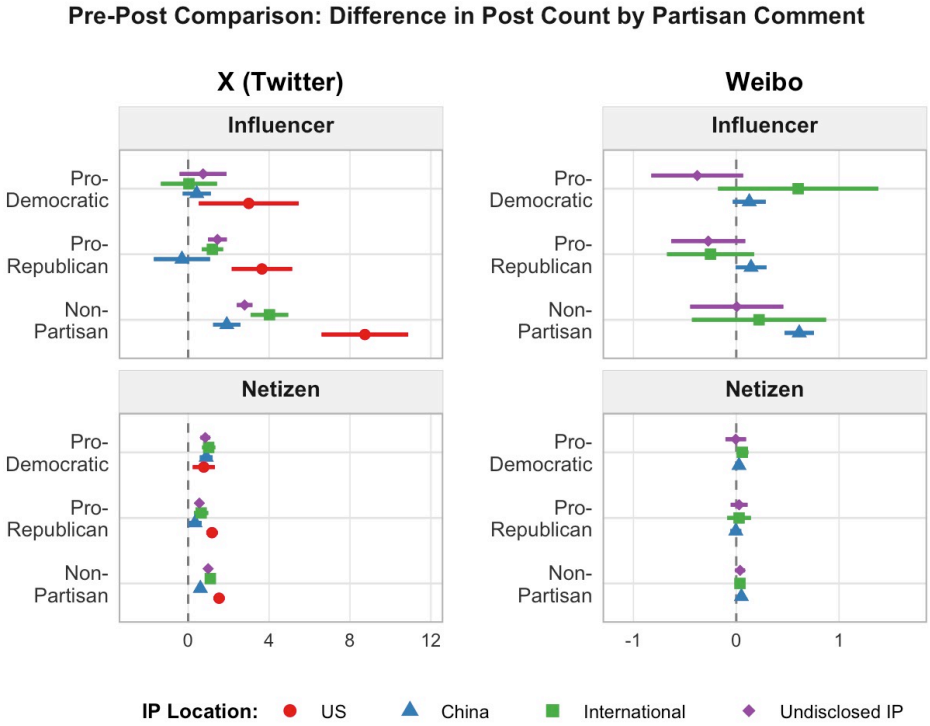
General users on X also exhibit heightened activity following the event, albeit with more modest effects than influencers. Irrespective of geographic location, the average general user produces approximately 1 to 2 additional message per content category

[6]Full regression estimates are in Appendix D.1.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved



(a) misinformation Content Creation by Time, User Type, and IP Locations



(b) Partisan Content Creation by Time, User Type, and IP Locations

Figure 3: Estimated Increase in Post Created by Content Creators After Assassination Attempt on Trump

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

post-event. While influencer behavior displays pronounced geographic heterogeneity, this variation diminishes considerably among general users. General users uniformly increase production of misinformation and partisan content relative to accurate and non-partisan content. This pattern suggests that general users respond to the post-event information environment in a manner that amplifies non-partisan and truthful content as well as biased and low-quality content.[7]

Weibo influencers exhibit less distinct geographic patterns in content creation. Mainland Chinese influencers demonstrate the most pronounced post-event response, driven largely by the presence of professional news agencies among this user category. This is probably because Weibo has many news-sharing accounts; many of them have millions of followers and some are endorsed by the state (King et al., 2017). These institutional accounts prioritize rapid news dissemination following major political events, resulting in predominantly accurate content. Specifically, mainland Chinese influencers on Weibo produce, on average, 0.5 additional mostly accurate messages, 0.2 additional partially misleading messages, and fewer than 0.1 additional mostly inaccurate messages compared to the pre-event baseline. In contrast, influencers from international locations or with undisclosed IP addresses show no significant changes in content creation following the assassination attempt.

General users on Weibo remain largely unresponsive to the event across all geographic categories. Unlike the patterns observed on X, Weibo's general user population—regardless of IP location—demonstrates no measurable increase in message creation following the assassination attempt. This platform-specific difference suggests that Weibo's ecosystem, dominated by institutional actors and subject to different content moderation policies, may not encourage spontaneous political engagement among ordinary users during international political crises.

Having established that influencers drive post-event content creation—with notable geographic heterogeneity across platforms—I now turn to examining how users engage with different types of content following the assassination attempt. This analysis shifts focus from content production to consumption patterns, revealing how political shocks alter user preferences for misinformation and partisan messaging. I employ a within-platform difference-in-differences design, estimating:

$$\text{Engagement}_{i,t} = \alpha + \beta_1 \text{Post}_t + \beta_2 \text{ContentType}_i + \gamma(\text{Post}_t \times \text{ContentType}_i) + \epsilon_{i,t} \quad (2)$$

[7]This amplification effect is confirmed by the positive and statistically significant coefficients for all time × content × netizen interactions in Table A.3 in Appendix D.1. On X, these coefficients range from 4.739 to 7.449, indicating that general users increase misinformation and partisan content production by approximately 5-7 additional posts post-event compared to their increase in accurate or non-partisan content. On Weibo, the corresponding effects are smaller but consistent (0.343 to 0.556 for misinformation, 0.415 to 0.463 for partisanship), suggesting this amplification pattern occurs across platforms but with different magnitudes.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

where $\text{Engagement}_{i,t}$ represents comments, likes, or retweets for message i at time t; Post_t indicates the post-event period; and ContentType_i captures either misinformation level or partisan orientation. The coefficient γ identifies the differential change in engagement for specific content types following the assassination attempt.

To ensure robust inference, I implement the design-based supervised learning (DSL) framework proposed by Egami et al. (2023), which addresses potential overfitting concerns in supervised learning applications to causal inference. All models use expert-coded variables as predictors with LLM-coded variables for out-of-sample prediction.

The assassination attempt produced differential engagement patterns across content types, revealing selective mobilization of user attention rather than uniform increases. Misinformation content showed no significant post-event changes in engagement, with interaction effects consistently negative but statistically insignificant across all engagement types. This contrasts with expectations that political shocks might increase appetite for unverified information. Partisan content revealed clear asymmetries when examined through binary indicators. Pro-Republican content, despite significant negative baseline effects (-4.903 for comments, -42.041 for likes, -8.326 for retweets), experienced large post-event increases (10.234 additional comments, 98.612 additional likes, 23.531 additional retweets). Anti-Democratic content followed similar patterns with even larger effects for likes (128.106 additional, $p<0.01$), while anti-Republican content showed minimal and inconsistent changes. The absence of pro-Democratic content from analysis due to insufficient sample size itself indicates the dominance of Republican-favorable discourse in Chinese social media responses to the event.

Affective polarization content demonstrated that ingroup affinity, rather than outgroup hostility, drove engagement increases. Ingroup liking content significantly increased retweet activity following the event (18.602, $p<0.1$), while outgroup hatred showed no significant effects across engagement types. This pattern suggests that political shocks amplify group cohesion behaviors rather than hostile responses. These findings indicate that political disruptions create selective attention patterns that align with pre-existing geopolitical tensions rather than generating indiscriminate information consumption. The assassination attempt reinforced existing partisan predispositions in Chinese social media discourse, mobilizing engagement with content supporting Republican figures and criticizing Democratic politicians while promoting ingroup solidarity behaviors.

Finally, I turn into the downstream behavioral effects of exposure to different volume of information across platforms. I exploit cross-platform variation in content exposure. X and Weibo offer distinct information environments—X exposes users to a substantially higher volume of information, as Figure A.2 and A.3 shows.[8] This variation enables a quasi-experimental design where X serves as the "high-exposure" treatment condition relative to Weibo's more regulated environment. I implement a cross-platform difference-

[8]See the two figures in Appendix C.1.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Table 4: Content Type and Online Engagement on Twitter (X)

	Information Type				Polarized Attitude	
	Misinfo	Pro-Rep	Anti-Rep	Anti-Dem	Outgroup	Ingroup
Panel A: Comment Models						
Intercept	5.743***	5.653***	5.647***	5.653***	5.457***	5.607***
	(0.442)	(0.363)	(0.363)	(0.363)	(0.414)	(0.383)
Post-Event	1.632***	1.385***	1.395***	1.384***	1.734**	1.495**
	(0.548)	(0.392)	(0.392)	(0.392)	(0.589)	(0.513)
Content	−0.348	−4.903***	1.153	−4.453***	2.213	1.184
	(0.659)	(0.551)	(5.426)	(0.751)	(2.784)	(3.242)
Post-Event × Content	−1.262	10.234**	2.049	12.593*	−4.443	−2.448
	(1.527)	(4.305)	(7.328)	(7.424)	(6.590)	(7.712)
Panel B: Like Models						
Intercept	48.097***	48.041***	48.040***	48.052***	47.288***	47.596***
	(9.695)	(8.164)	(8.166)	(8.166)	(9.024)	(8.516)
Post-Event	−0.248	−1.720	−1.640	−1.741	−1.982	−3.691
	(9.924)	(8.260)	(8.262)	(8.261)	(9.291)	(8.764)
Content	−0.352	−42.041***	−32.440*	−42.652***	8.262	11.530
	(8.011)	(9.262)	(15.784)	(9.012)	(20.872)	(26.684)
Post-Event × Content	−6.720	98.612***	36.040†	128.106**	8.615	45.823
	(11.258)	(29.149)	(24.700)	(42.591)	(34.248)	(44.774)
Panel C: Retweet Models						
Intercept	6.946***	8.572***	8.566***	8.567***	8.340***	8.459***
	(1.455)	(0.883)	(0.883)	(0.883)	(0.986)	(0.927)
Post-Event	1.748	0.571	0.602	0.576	0.776	−0.228
	(1.546)	(0.919)	(0.920)	(0.920)	(1.162)	(1.062)
Content	5.931†	−8.326***	−5.566*	−5.566*	2.543	2.853
	(4.342)	(0.909)	(2.825)	(2.825)	(4.270)	(6.125)
Post-Event × Content	−3.614	23.531**	4.227	27.482**	−1.773	18.602†
	(4.868)	(8.425)	(4.222)	(11.696)	(9.477)	(11.665)

Notes: All models estimated with the DSL procedure (Egami et al. 2023). Misinformation coded 0–2 (accurate to misleading); Partisan and affective polarization coded in binary indicators. Pro-Democratic models could not be estimated due to insufficient coverage of pro-Democratic posts in the expert-labeled sample. Labeled observations: $n = 35{,}026$, with $n = 500$ expert-labeled posts. Standard errors in parentheses. $^{\dagger}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

in-differences specification:

$$Y_{i,p,t} = \alpha + \beta_1(\text{Post}_t \times \text{X}_p) + \beta_2\text{Post}_t + \beta_3\text{X}_p + \epsilon_{i,p,t} \quad (3)$$

where $Y_{i,p,t}$ represents behavioral outcomes for user i on platform p at time t. I examine six outcomes capturing both quantity and quality of user-generated content: total messages created, a count outcome for misinformation, a count out come for partisan content (pro-Republican and pro-Democrat separately), a measure of affective polarization, and attack language. The coefficient β_1 identifies the average treatment effect on the treated (ATT)—the differential behavioral change among X users following the assassination attempt, relative to the counterfactual trend represented by Weibo users.

Table 5: Difference-in-Differences Analysis: Platform Effects on Content Production

	Outcome Variables					
	Post Count (1)	Affective Polarization (2)	Misinformation Posts (3)	Pro-Republican Posts (4)	Pro-Democratic Posts (5)	Attack Posts (6)
Baseline Estimates						
Intercept (Weibo baseline)	1.187***	0.066***	0.086***	0.125***	0.077***	0.109***
	(0.014)	(0.005)	(0.005)	(0.007)	(0.005)	(0.006)
Post-Event	0.053***	0.035***	0.057***	0.018*	0.048***	0.022***
	(0.013)	(0.006)	(0.006)	(0.008)	(0.006)	(0.006)
Using X/Twitter	0.997***	0.370***	0.376***	0.496***	0.449***	0.481***
	(0.062)	(0.018)	(0.019)	(0.026)	(0.036)	(0.021)
DiD Estimates						
Post-Event × Using X	1.329***	0.183***	0.237***	0.285***	0.115***	0.175***
	(0.054)	(0.019)	(0.018)	(0.024)	(0.028)	(0.020)
Model Statistics						
Observations	50,621	50,621	50,621	50,621	50,621	50,621
R^2	0.021	0.014	0.014	0.015	0.006	0.014
RMSE	7.21	2.09	2.38	2.82	3.21	2.46

Notes: Robust standard errors clustered by username in parentheses. All models use OLS estimation. The dependent variables measure user content production in the pre- and post-event periods. Post-Event indicates the period following the triggering event. Using X/Twitter indicates users primarily active on X/Twitter versus Weibo. The interaction term captures the differential effect of the event for X users relative to Weibo users. Misinformation posts are counted when misinformation level is ≥ 0.5. Affective polarization measures the sum of ingroup favoritism and outgroup hostility. Intercept represents baseline levels for Weibo users in the pre-event period. Significance levels: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 5 shows downstream behavioral impacts from platform-mediated exposure. The difference-in-differences estimates show that X users increase their posting frequency by 1.33 additional posts on average and demonstrate heightened affective polarization, with 0.18 additional expressions of ingroup favoritism or outgroup hostility. Also, the informational quality of their content deteriorates: users generate 0.24 additional misinformation messages—a substantial increase given baseline rates of 0.09 posts.

The partisan composition of user-generated content also displays notable shifts. X users produce 0.29 additional pro-Republican messages and 0.11 additional pro-Democratic messages, indicating both increased political expression and asymmetric partisan mobi-

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

lization strongly favoring Republican narratives. The civility of discourse similarly declines, with users generating 0.18 additional messages containing attack language. These effect magnitudes demonstrate that exposure to am information environment richer in partisan discussion shapes user behavior beyond passive consumption.

The causal interpretation of the difference-in-differences estimates requires the parallel trends assumption—that X and Weibo users would have followed similar behavioral trajectories absent the assassination attempt. This assumption faces a particular challenge in social media contexts where users self-select into platforms based on pre-existing political attitudes and information preferences. To address this concern and test for differential pre-trends, I implement a dynamic two-way fixed effects (TWFE) specification:

$$Y_{i,t} = \alpha_i + \delta_t + \sum_{k=-10}^{14} \beta_k \cdot \mathbf{1}[\mathrm{X}_i \times \mathrm{Days}_{i,t} = k] + \epsilon_{i,t} \tag{4}$$

where $Y_{i,t}$ represents political engagement outcomes for user i at time t; α_i captures user fixed effects that absorb time-invariant characteristics including platform selection; δ_t denotes day fixed effects controlling for common temporal shocks; and $\mathbf{1}[\mathrm{X}_i \times \mathrm{Days}_{i,t} = k]$ indicates whether user i on platform X is observed k days from the assassination attempt. The coefficients $\{\beta_k\}_{k=-10}^{14}$ trace out the dynamic treatment effects, with β_{-1} normalized to zero as the reference period.

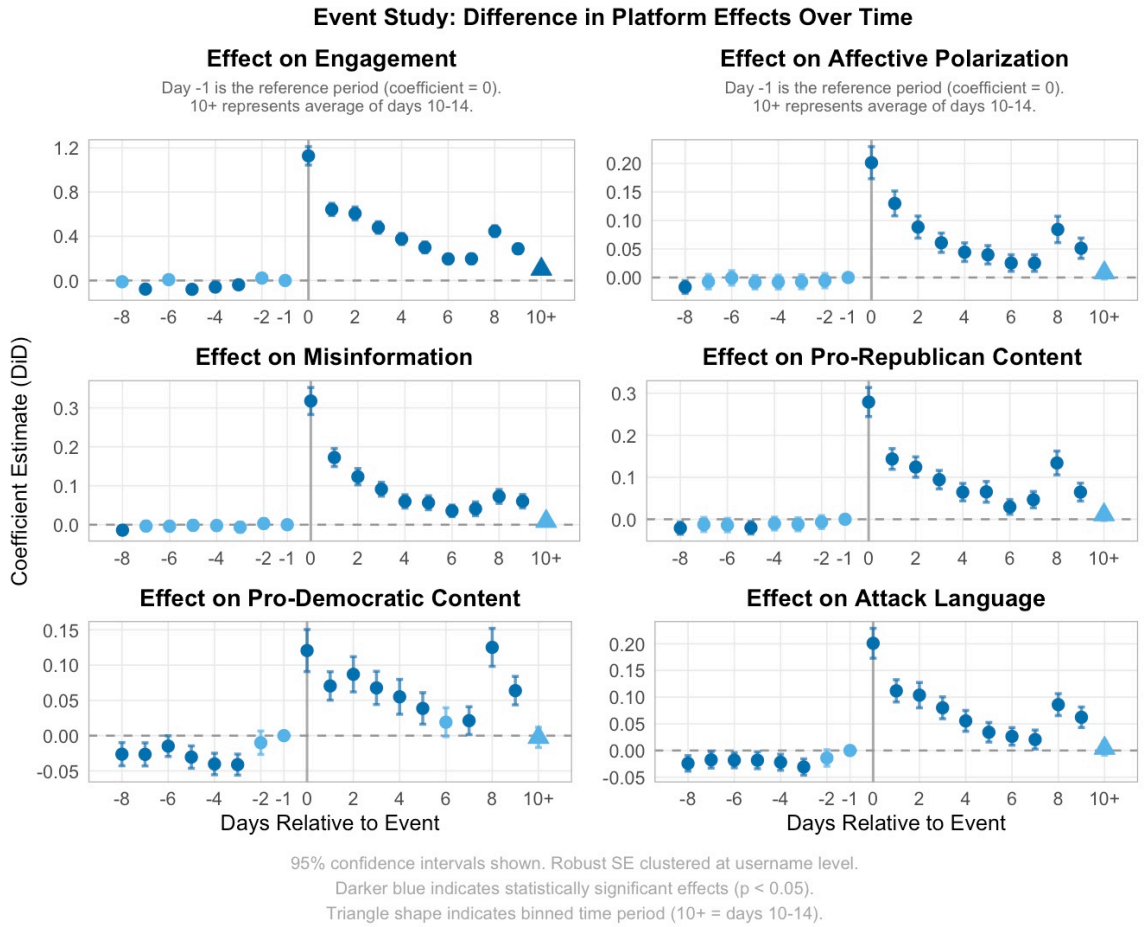
This event-study design serves two purposes. First, the pre-treatment coefficients $\{\beta_k\}_{k=-10}^{-2}$ test for differential trends that would violate the identifying assumption. Statistically insignificant pre-treatment coefficients support the parallel trends assumption, suggesting that behavioral differences emerge only after the assassination attempt rather than reflecting pre-existing trajectories. Second, the post-treatment coefficients $\{\beta_k\}_{k=0}^{14}$ reveal the temporal dynamics of behavioral contagion, showing whether effects persist, amplify, or decay over time.

Figure 4 presents the event study estimates with 95% confidence intervals. The pre-treatment coefficients ($k < 0$) remain mostly statistically insignificant across three outcome variables, including affective polarization, misinformation, and pro-Republican inclinations. For the outcome variables that show a pre-treatment differences (engagement, pro-Democratic content, attack language use), there is no systematic trends. The combined evidence confirms that parallel assumption won't become a threat to the DiD estimations: X and Weibo users followed similar behavioral trajectories before the assassination attempt, alleviating concerns that platform-specific trends drive the results. This validation supports interpreting the post-event differences as causal effects of differential information exposure rather than artifacts of user selection.

The post-treatment dynamics ($k \geq 0$) reveal immediate and persistent behavioral changes. All outcomes show significant positive effects beginning on the event date, with

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Figure 4: Platform Exposure Effects on User Behavior: Event Study Analysis



Notes: The plots show coefficients from dynamic two-way fixed effects (TWFE) models estimated via OLS, with user and day fixed effects included. The analysis uses a balanced daily panel of users from X/Twitter (treatment group) and Weibo (control group), restricted to a cohort of users active before the event. A robustness check using a larger panel unfiltered by activeness shows similar results with better statistical significance. Day -1 serves as the reference period, with its coefficient normalized to zero. The period from Day +10 to Day +14 is collapsed into a single "10+" estimate representing the average effect in that window. On day 8 (July 21) after the assassination, President Biden withdrew his candidacy from the next presidential election. Shaded areas represent 95% confidence intervals based on standard errors clustered at the user level. Given the non-staggered treatment timing, the TWFE estimator provides robust causal estimates.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

these effects remaining elevated at least ten days following the assassination attempt. Message creation, affective polarization, misinformation production, pro-Republican content, and attack language maintain consistently positive coefficients until day 10, suggesting that exposure to X's information environment induces durable rather than transient behavioral changes. The persistence of these effects is particularly noteworthy—while behavioral responses to news events typically attenuate as initial shocks fade, the sustained elevation indicates that exposure to an information environment richer in partisan conversation may shape users' information-sharing norms at least during the event's window. Pro-Democratic content represents the sole exception, showing positive but statistically insignificant effects on an earlier date, consistent with the asymmetric partisan message production documented earlier.

An important contextual consideration emerges on day 8 (July 21), when President Biden withdrew from the presidential race. While this event generated renewed discussion activity, it does not confound our results for two reasons. First, much post-withdrawal discussion remained assassination-related, as users attributed Biden's exit to Democrats' strategic adjustments following the assassination attempt.[9] Second, the event study design clearly delineates the magnitude and duration of X usage effects, showing that the primary behavioral changes preceded and persisted through this secondary event.

Figure 5 disaggregates the event-study results by users' geographic location, revealing substantial heterogeneity in behavioral responses to platform exposure.[10] U.S.-based Chinese-speaking users exhibit the strongest treatment effects across all dimensions. Following the assassination attempt, they show pronounced and persistent increases in posting frequency, misinformation dissemination, and partisan content creation. These effects remain statistically significant throughout the 14-day post-event window, suggesting that geographic proximity to American politics amplifies susceptibility to information environment effects.

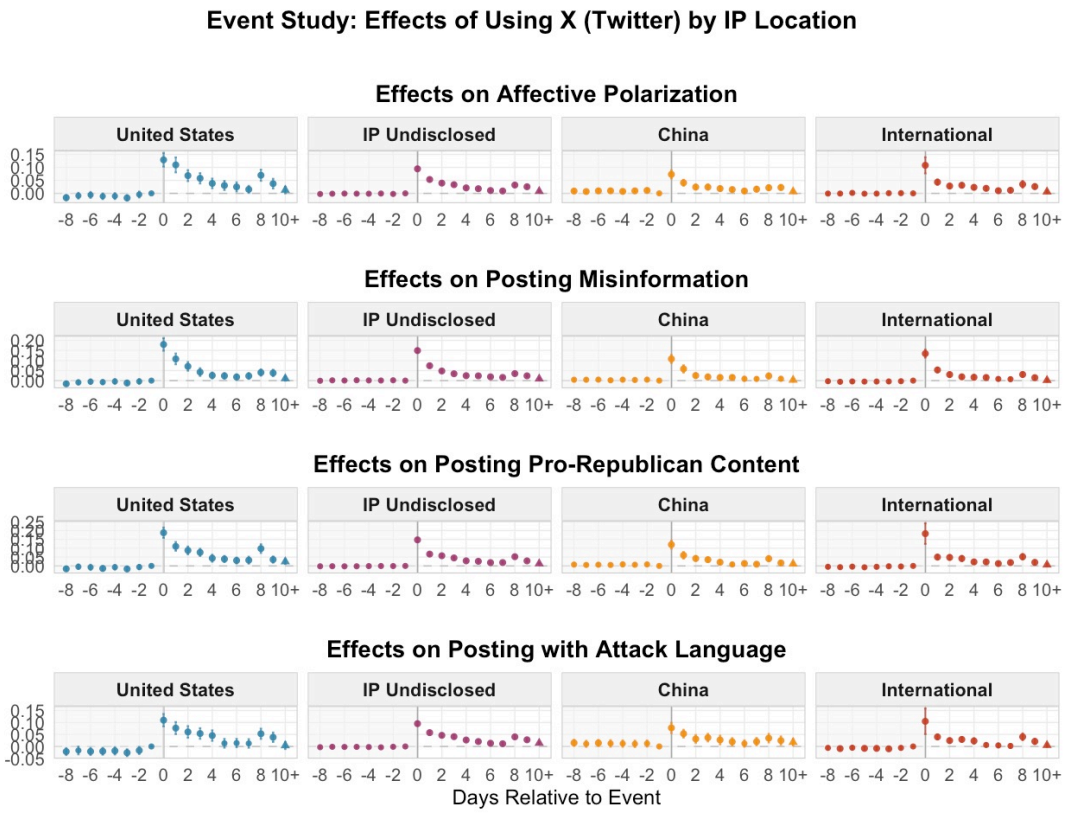
It's interesting to note that both International Chinese-speaking users (those outside both China and the U.S.) and mainland Chinese users display significant and consistent increase in partisan-like downstream online behavior, with sustained increases in post generation that contains affective polarization attitudes, partisan inclinations, misinformation sharing, and attack language. This effect lasts until at least Day 6, although the effect sizes are smaller compared to those demonstrated by U.S.-based Chinese-language users. This pattern indicates that the behavioral contagion extends beyond users with direct stakes in U.S. politics, affecting the broader Chinese-speaking diaspora engaged with American political events.

[9] Appendix C.2 presents comparative analysis of discussion volume for posts containing "assassination" and "Trump" keywords relative to overall discussion after day 8.

[10] Full regression estimates are in Appendix D.2.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Figure 5: Platform Exposure Effects by Geographic Location: Event Study Analysis



Notes: This subgroup analysis applies a pre-event user activity filter to ensure data quality, the same way as the full sample event-study analysis does. Robustness checks without this filter preserve larger sample sizes and yield similar results with greater statistical significance.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Study 2

Study 2 examines selective exposure behavior surrounding the 2024 U.S. presidential election using data from Rednote. I operationalize "selective exposure" as users' tendency to engage with content by posting comments that align with the ideological direction of the original post. The dataset comprises approximately 9,000 posts and 200,000 associated comments collected during a four-week window spanning two weeks before and after the November 5, 2024 election.[11] This temporal framework captures both anticipatory political discourse and post-election reactions, enabling analysis of how electoral uncertainty and its resolution shape information-sharing behaviors.

Rednote provides a unique analytical advantage through its mandatory IP geolocation disclosure. Unlike X's self-reported locations, Rednote automatically captures and displays users' geographic locations in compliance with China's Cybersecurity Law provisions requiring real-identity verification on social media platforms.[12] This feature eliminates concerns about strategic location misrepresentation and enables precise geographic stratification of behavioral responses to the U.S. election results.

First, I examine information production on Rednote before and after election. I fit the following regression model to identify changes in content patterns following the election results:

$$Y_{i,t} = \alpha + \beta_1 \text{PostElection}_t + \beta_2 \text{Location}_i + \beta_3 (\text{PostElection}_t \times \text{Location}_i) + \epsilon_{i,t} \quad (5)$$

where $Y_{i,t}$ represents content characteristics for user i at time t, including posting volume, misinformation levels, and partisan orientation. The binary indicator PostElection_t equals 1 for content generated after November 5, 2024, capturing the period when the election has concluded and election results have became definitive. The coefficient β_3 identifies differential responses across user locations to the period of content creation.

Figure 6's Panel A examines post-level characteristics, while Panel B analyzes comment-level behaviors across geographic strata.[13] The results reveal asymmetric responses between content creators and commenters. Following Trump's electoral victory, post creation declines significantly across all geographic locations, suggesting that political uncertainty before election resolution perhaps drives content production. However, the ideological composition of posts shifts markedly—creators generate significantly more pro-Republican content post-election, indicating that electoral results influence the partisan valence of discourse even as overall volume decreases.

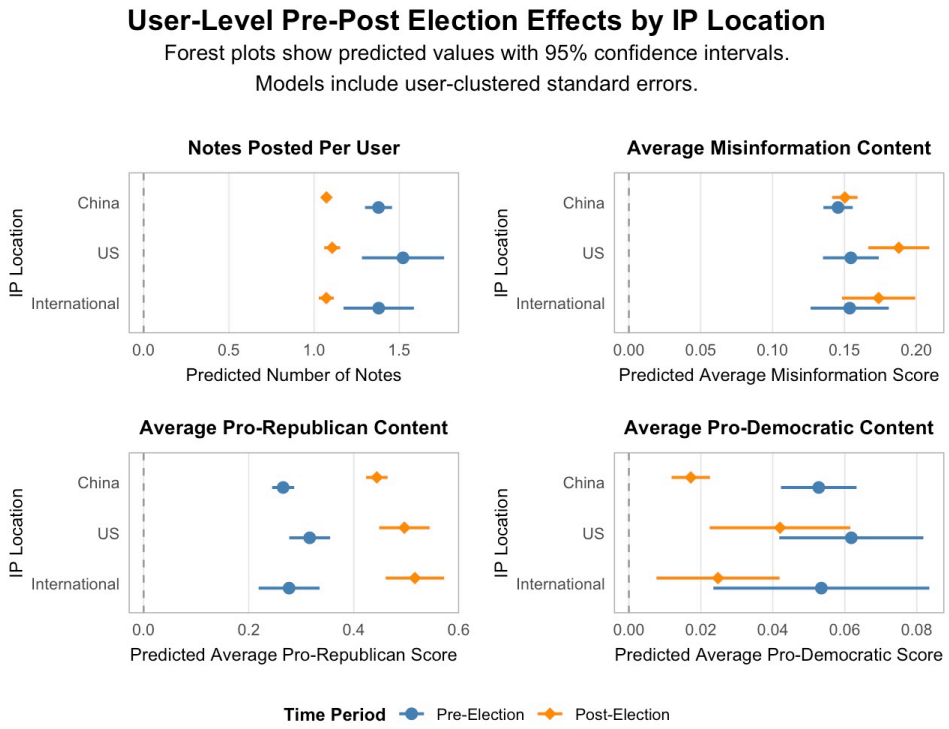
Comment-level patterns exhibit distinct dynamics. While the volume of comments per

[11] See a time-series plot of how post and comment density change throughout the window in Appendix C.3.

[12] The Cybersecurity Law of the People's Republic of China, implemented June 1, 2017, mandates that network operators verify users' real identities and maintain accurate registration information.

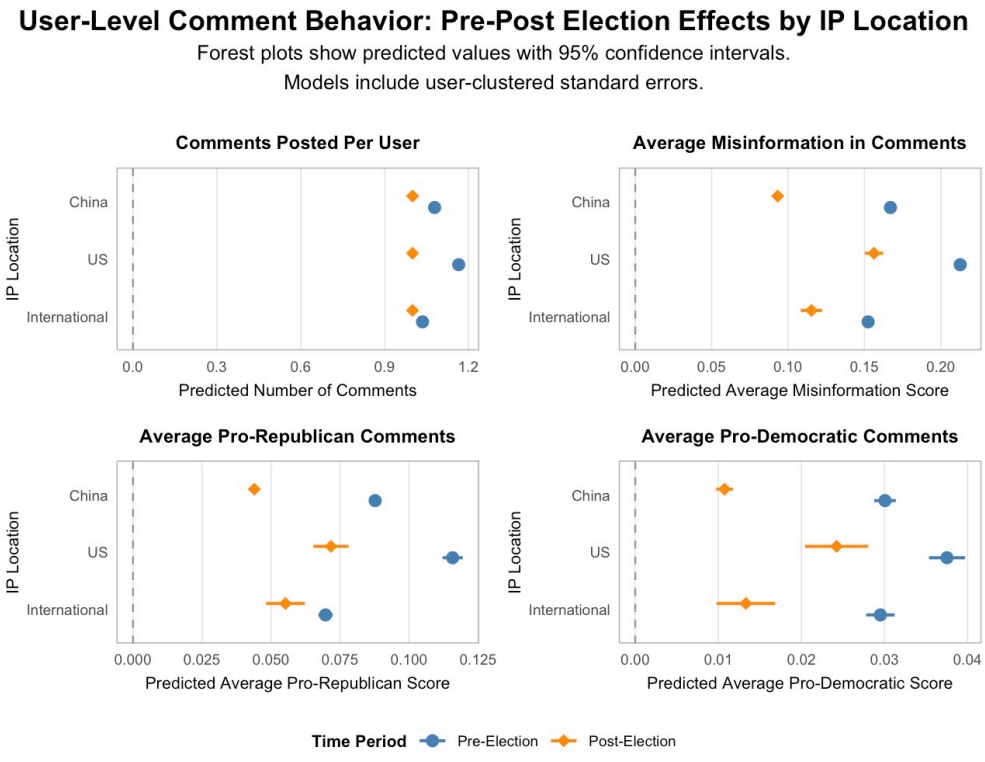
[13] Full regression estimates are in Appendix D.3.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved



Note: Analysis based on all users who either posted before or after the 2024 US Election (Nov 5).

(a) Post Characteristics by Post-Election and IP Addresses



Note: Analysis based on users who either commented before or after the 2024 US Election (Nov 5).

(b) Comment Characteristics by Post-Election and IP Addresses

Figure 6: Average Content Characteristics During U.S. Presidential Election on Rednote

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

user remains stable across the pre- and post-election periods, the informational quality of comments improves substantially. Both misinformation levels and partisan content in comments decline significantly after November 5, suggesting that electoral resolution is negatively correlated with speculative and politically charged user responses.

Geographic heterogeneity persists throughout both periods. U.S.-based Chinese speakers consistently generate higher levels of misinformation and partisan content in their comments compared to users from mainland China or other countries. This elevated baseline remains stable across the electoral transition, indicating that geographic proximity to U.S. politics creates enduring differences in how Chinese-speaking users engage with American electoral events.

Second, I examine selective exposure behavior by analyzing how characteristics of original posts correlate with the properties of subsequent user comments. This analysis leverages Rednote's hierarchical data structure, where unique identifiers link each comment to its parent post, enabling precise measurement of exposure effects. However, these estimates should be interpreted as correlational rather than causal, since users encounter content through algorithmic curation based on their long-term consumption patterns rather than random exposure (Green et al., 2025). Nevertheless, if I observe positive correlations between post characteristics and comment characteristics—where users preferentially engage with ideologically aligned content—this pattern would support my selective exposure hypothesis.

I estimate two complementary specifications to identify both temporal and geographic heterogeneity of these correlations:

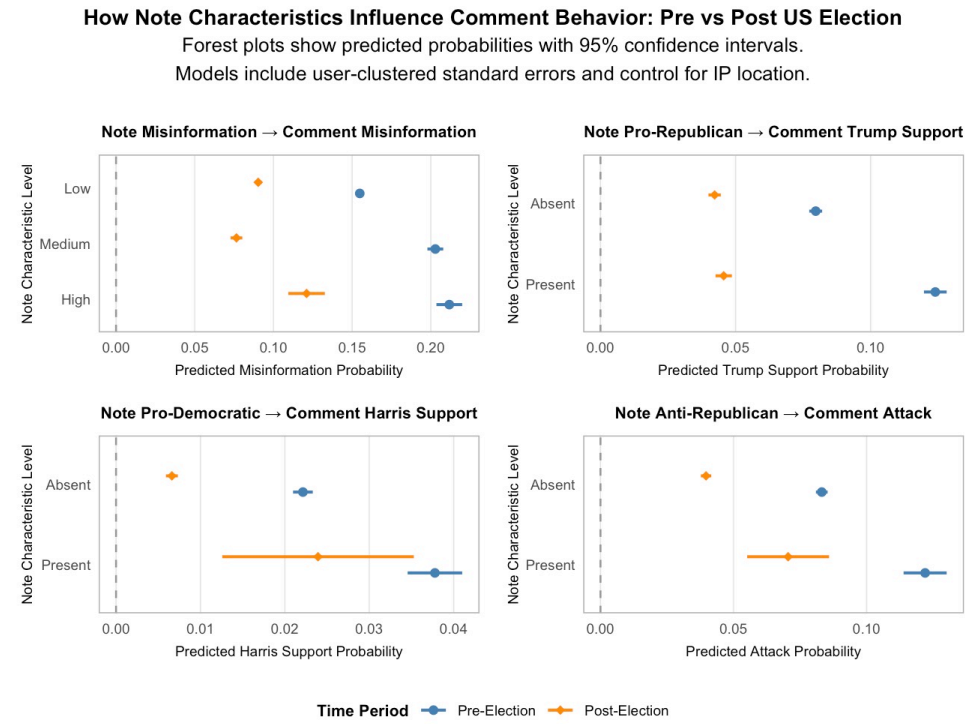
$$\text{Comment}_{ij} = \alpha + \beta_1 \text{PostChar}_j + \beta_2 \text{PostElection}_t + \beta_3 (\text{PostChar}_j \times \text{PostElection}_t) + \epsilon_{ij} \tag{6}$$

$$\text{Comment}_{ij} = \alpha + \beta_1 \text{PostChar}_j + \beta_2 \text{Location}_i + \beta_3 (\text{PostChar}_j \times \text{Location}_i) + \epsilon_{ij} \tag{7}$$

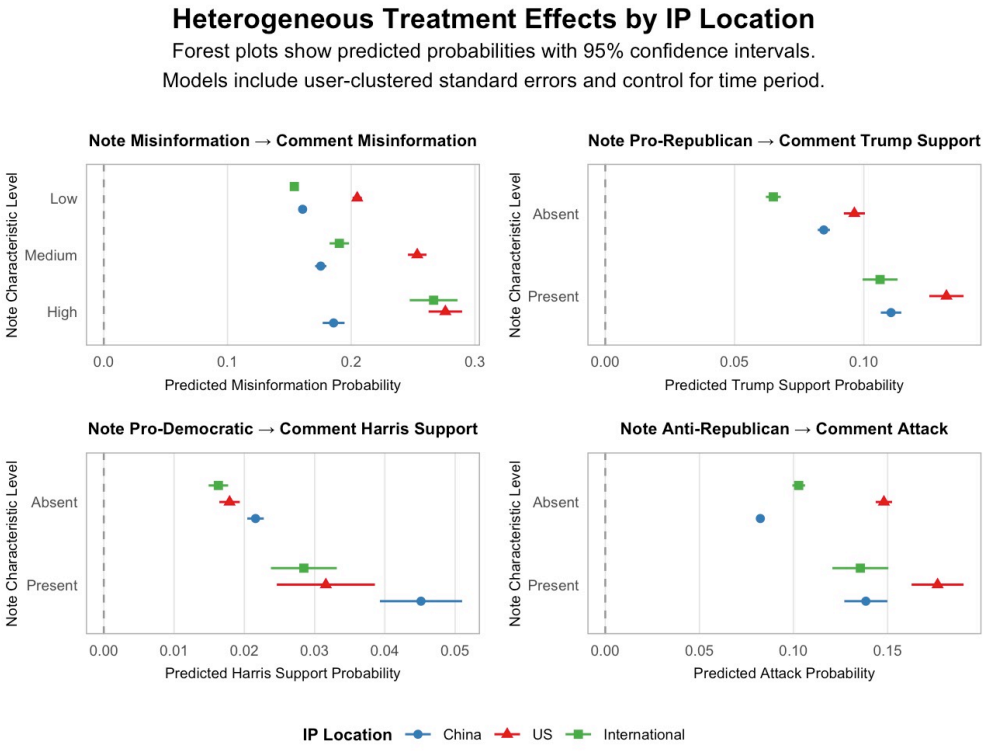
where Comment_{ij} captures the informational characteristics of a comment by user i responding to post j, including misinformation level, partisan orientation, and presence of attack language. PostChar_j represents the corresponding characteristics of the parent post, providing the measure of information exposure. In equation (6), PostElection_t indicates whether the comment was created after November 5, while in equation (7), Location_i denotes the commenter's geographic location.

The coefficient β_1 identifies the baseline effect—the degree to which exposure to posts with specific characteristics shapes the informational quality of user responses. The interaction coefficients β_3 in each specification capture heterogeneity: equation (6) tests whether electoral resolution moderates the correlations, while equation (7) examines whether users' geographic contexts influence their responsiveness to misleading or partisan content.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved



(a) How Post Characteristics Influence Commenting Behavior by Pre/Post-Election



(b) Comment Characteristics by Post-Election and IP Addresses

Figure 7: How Post Characteristics Influence Commenting Behavior by IP Locations

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Figure 7 presents the estimates.[14] Panel A reveals how electoral resolution affects different transmission patterns of misinformation and partisan content. Prior to November 5, exposure to posts containing misinformation is significantly more correlated with the probability of generating misleading comments compared to the post-election period. After the election, exposure to moderately misleading posts is no longer correlated with increased misinformation production in comment sections. Even highly misleading posts show only modest correlations with increased comment misinformation. Moreover, the baseline propensity to post misleading comments declines significantly after November 5, suggesting that electoral resolution reduces both susceptibility to misinformation and its organic generation.

Partisan posts exhibit even more pronounced temporal variation. Before the election, exposure to pro-Republican posts is correlated with a significant increase in the density of pro-Trump comments—evidence of strong partisan echo chambers during the campaign period. Post-election, however, pro-Republican posts are no longer correlated with increased density of pro-Trump comments. The overall density of pro-Trump comments similarly declines after November 5. Pro-Democrat content shows distinct dynamics. While exposure to pro-Democratic posts is correlated with more pro-Democratic comments across both periods, the baseline frequency of pro-Harris comments drops significantly post-election. This provides further evidence for the existence of partisan echo chambers on Rednote.

Panel B examines geographic heterogeneity in behavioral contagion from posts to comments. The results reveal that while baseline commenting behaviors vary substantially across geographic locations, users from all geographical locations exhibit selective exposure behavior. U.S.-based Chinese speakers exhibit systematically different baseline characteristics compared to their mainland Chinese counterparts. They demonstrate higher propensities for sharing misinformation, expressing pro-Trump sentiment, and employing attack language. Despite these substantial baseline differences, the correlations show remarkable uniformity across locations. Exposure to misinformation is correlated with increased probability of generating misleading comments by similar magnitudes for users in the U.S., mainland China, and other international locations. This geographic invariance extends to the correlation between seeing partisan posts and producing partisan comments: pro-Republican posts are correlated with increased pro-Trump comments, and pro-Democrat posts are correlated with increased pro-Harris comments by comparable amounts regardless of user location. This suggests that selective exposure behavior can extend from its original population to geographically and linguistically remote groups, albeit with weaker magnitude.

[14]Full regression estimates are in Appendix D.4.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Conclusion

This paper began with Chinese-speaking social media users thousands of miles from Pennsylvania debating bullet trajectories and divine intervention following the Trump assassination attempt. Their engagement with U.S. political events—despite lacking voting rights or direct stakes in electoral outcomes—exemplifies a broader phenomenon: polarized political discourse spreads through digital networks regardless of geographic or linguistic boundaries.

The empirical findings reveal how political shocks cascade through information ecosystems. When major events occur, influencers flood platforms with more misleading and partisan content. Users preferentially engage with this low-quality information, amplifying its reach. Exposure to misinformation and partisan content then shapes user behavior, inducing them to produce similar content themselves.

Two aspects of these findings deserve emphasis. First, the effects operate similarly across different political contexts. Whether users reside in authoritarian China, the U.S., or the diaspora, exposure to misleading content increases their propensity to spread misinformation by comparable magnitudes, although Chinese speakers "closer" to the election show greater susceptibility. This suggests that susceptibility to partisan discourse may reflect psychological features rather than merely cultural or institutional contexts.

Second, platform architecture plays an important role in facilitating or constraining contagion. X's minimal content moderation creates an environment where misinformation thrives, while Weibo's stricter controls better limit—though do not eliminate—the spread of false content. The cross-platform variation in user behavior demonstrates that platform policies causally affect information quality, not merely through direct content removal but by shaping user norms and expectations.

Interestingly, geographic and temporal distance provides only limited protection against the contagious effects of misinformation. Many Chinese speakers engage actively with U.S. political events despite lacking electoral stakes, accessing information through Chinese content produced by a mixture of honest and misleading sources. Their engagement patterns mirror those of U.S. audiences, as they map their unobserved political identities onto the mega-identity of U.S. partisanship and translate this mapping into actions. These findings have immediate implications for platform governance. While platforms enable unprecedented cross-cultural exchange, they also facilitate the transmission of political pathologies across borders.

Several limitations merit acknowledgment. First, misinformation and partisan content are difficult to label accurately, with the most up-to-date Large Language Models (e.g., ChatGPT-4.1) achieving 80-90 percent accuracy and previous generations' models (e.g., ChatGPT-4o-mini) achieving 60-70 percent. Future research with greater computational resources should employ more advanced models for content classification. Second,

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

this largely observational study cannot fully overcome selection bias. The sample includes more politically engaged users than the general population, as inattentive users are unlikely to post or comment on political events (Mukerjee et al., 2022). Therefore, findings may reflect only the subset of politically engaged Chinese-language speakers on social media. Future experimental research with random representative samples would strengthen causal interpretations.

Future research could explore the mechanisms driving this transnational spread of partisan messages. While my data demonstrates that this phenomenon exists, understanding the psychological processes that motivate Chinese-language users to engage with distant politics—and clarifying their underlying motivations—would provide valuable insights. Additionally, future studies could examine whether similar dynamics occur in other linguistic communities. Do Spanish speakers in Latin America show comparable patterns when engaging with European politics? How do English-speaking audiences process political shocks from non-Anglophone countries? Understanding the full scope of transnational political influence is essential as democracies worldwide grapple with globally networked information ecosystems.

The transformation of distant witnesses into active participants in foreign political discourse represents an important shift in how political information circulates globally. As digital platforms continue to collapse geographic and linguistic boundaries, the quality of democratic discourse in any nation increasingly depends on information dynamics everywhere. The contagion of distant politics is not merely a curiosity of our interconnected age—it is a central challenge for maintaining epistemic communities capable of supporting democratic governance in the twenty-first century.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

References

Abrajano, Marisa, Garcia, Marianna, Pope, Aaron, Vidigal, Robert, Tucker, Joshua A, and Nagler, Jonathan (Nov. 2024). How reliance on Spanish-language social media predicts beliefs in false political narratives amongst Latinos. *PNAS Nexus* 3.11, pgae442.

Adair, Bill (Feb. 2018). The principles of the Truth-O-Meter: PolitiFact's methodology for independent fact-checking. *PolitiFact*.

Allen, J., Howland, B., Mobius, M., Rothschild, D., and Watts, Duncan J. (2024). Quantifying the role of chance in collective behavior. *Proceedings of the National Academy of Sciences* 121.5, e2308063121.

Bail, Christopher A., Argyle, Lisa P., Brown, Taylor W., Bumpus, John P., Chen, Haohan, Hunzaker, M. B. Falco, Lee, Jaemin, Mann, Marcus, Merhout, Friedolin, and Volfovsky, Alexander (2018). Exposure to Opposing Views on Social Media Can Increase Political Polarization. *Proceedings of the National Academy of Sciences* 115.37, 9216–9221.

Barberá, Pablo, Jost, John T., Nagler, Jonathan, Tucker, Joshua A., and Bonneau, Richard (2015). Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychological Science* 26.10, 1531–1542.

Bartels, Larry M. (June 2002). Beyond the running tally: partisan bias in political perceptions. *Political Behavior* 24.2, 117–150.

Birkland, Thomas A. (1998). Focusing Events, Mobilization, and Agenda Setting. *Journal of Public Policy* 18.1, 53–74.

Bossetta, Michael (2018). The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 U.S. Election. *Journalism and Mass Communication Quarterly* 95.2, 471–496.

Chen, Canyu and Shu, Kai (2024). Combating misinformation in the age of LLMs: Opportunities and challenges. *AI Magazine* 45.3, 354–368.

Chouliaraki, Lilie (2016). Human rights in an age of distant witnesses: remixed lives, reincarnated images and live-streamed co-presence. *The International Journal of Human Rights* 20.7, 929–944.

Egami, Naoki, Hinck, Musashi, Stewart, Brandon M., and Wei, Hanying (2023). Using Imperfect Surrogates for Downstream Inference: Design-Based Supervised Learning for Social Science Applications of Large Language Models. *Advances in Neural Information Processing Systems* 36, 68589–68601.

Elkins, Zachary and Simmons, Beth (2005). On waves, clusters, and diffusion: A conceptual framework. *The Annals of the American Academy of Political and Social Science* 598.1, 33–51.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Falkenberg, M., Zollo, F., Quattrociocchi, W., et al. (2024). Patterns of Partisan Toxicity and Engagement Reveal the Common Structure of Online Political Communication Across Countries. *Nature Communications* 15, 9560.

Gentzkow, Matthew and Shapiro, Jesse M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics* 126.4, 1799–1839.

Gou, Zhibin, Shao, Zhihong, Gong, Yeyun, Yang, Yujiu, Huang, Minlie, Duan, Nan, Chen, Weizhu, et al. (2023). ToRA: A tool-integrated reasoning agent for mathematical problem solving. *arXiv preprint arXiv:2309.17452*.

Green, Donald, Palmquist, Bradley, and Schickler, Eric (2002). *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. Yale University Press, 288.

Green, Jon, McCabe, Stefan, Shugars, Sarah, Chwe, Hanyu, Horgan, Luke, Cao, Shuyang, and Lazer, David (2025). Curation Bubbles. *American Political Science Review*, FirstView.

Grinberg, Nir, Joseph, Kenneth, Friedland, Lisa, Swire-Thompson, Briony, and Lazer, David (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science* 363.6425, 374–378.

Guess, Andrew and Coppock, Alexander (2018). Does counter-attitudinal information cause backlash? Results from three large survey experiments. *British Journal of Political Science*, 1–19.

Holliday, D. E., Lelkes, Y., and Westwood, S. J. (2024). The July 2024 Trump assassination attempt was followed by lower in-group support for partisan violence and increased group unity. *Proceedings of the National Academy of Sciences* 121.49, e2414689121.

Hopkins, Daniel J., Lelkes, Yphtach, and Wolken, Samuel (2024). The Rise of and Demand for Identity-Oriented Media Coverage. *American Journal of Political Science*.

Huang, Jie and Chang, Kevin Chen-Chuan (2022). Towards reasoning in large language models: A survey. *arXiv preprint arXiv:2212.10403*.

Huang, Tianyi et al. (2024). Unmasking digital falsehoods: A comparative analysis of LLM-based misinformation detection strategies. *arXiv preprint arXiv:2503.00724*.

Huddy, Leonie, Mason, Lilliana, and Aarøe, Lene (2015). Expressive Partisanship: Campaign Involvement, Political Emotion, and Partisan Identity. *American Political Science Review* 109.1, 1–17.

Iyengar, Shanto, Lelkes, Yphtach, Levendusky, Matthew, Malhotra, Neil, and Westwood, Sean J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science* 22, 129–146.

Iyengar, Shanto and Westwood, Sean J. (2015). Fear and loathing across party lines: new evidence on group polarization. *American Journal of Political Science* 59.3, 690–707.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

King, Gary, Pan, Jennifer, and Roberts, Margaret E. (2013). How Censorship in China Allows Government Criticism but Silences Collective Expression. *American Political Science Review* 107.2, 326–343.

King, Gary, Pan, Jennifer, and Roberts, Margaret E. (2017). How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument. *American Political Science Review* 111.3. Publisher's version available at `https://tinyurl.com/ycvo9zog`, 484–501.

Kobayashi, Tetsuro, Zhang, Zhifan, and Liu, Ling (2024). Is Partisan Selective Exposure an American Peculiarity? A Comparative Study of News Browsing Behaviors in the United States, Japan, and Hong Kong. *Communication Research* 0.0.

Kojima, Takeshi, Gu, Shixiang Shane, Reid, Machel, Matsuo, Yutaka, and Iwasawa, Yusuke (2022). Large language models are zero-shot reasoners. *Advances in Neural Information Processing Systems* 35, 22199–22213.

Krosnick, Jon A. and Petty, Richard E. (1995). Attitude Strength: Antecedents and Consequences.

Li, Junyi, Tang, Tianyi, Zhao, Wayne Xin, and Wen, Ji-Rong (2022). Retrieval-augmented generation for knowledge-intensive NLP tasks. *arXiv preprint arXiv:2204.07930*.

Little, Andrew T. (2025). How to distinguish motivated reasoning from Bayesian updating. *Political Behavior*. Advance online publication.

Lu, Yingdan, Pan, Jennifer, Xu, Xu, and Xu, Yiqing (2025). Decentralized Propaganda in the Era of Digital Media: The Massive Presence of the Chinese State on Douyin. *SSRN Electronic Journal*.

Mason, Lilliana (2015). 'I Disrespectfully Agree' : The Differential Effects of Partisan Sorting on Social and Issue Polarization. *American Journal of Political Science* 59.1, 128–145.

Mattozzi, Andrea, Nocito, Samuel, and Sobbrio, Francesco (2022). Fact-checking politicians. *CESifo Working Paper No. 10122*.

Mukerjee, Subhayan, Jaidka, Kokil, and Lelkes, Yphtach (2022). The Political Landscape of the US Twitterverse. *Political Communication*, 1–31.

Nyhan, Brendan (2020). Facts and myths about misperceptions. *Journal of Economic Perspectives* 34.3, 220–236.

Nyhan, Brendan, Settle, Jaime, Thorson, Emily, Wojcieszak, Magdalena, Barberá, Pablo, Chen, Annie Y., Allcott, Hunt, Bail, Christopher, Gentzkow, Matthew, Guess, Andrew, Munger, Kevin, Rae, Alejandra, and Tucker, Joshua A. (2023). Like-minded sources on Facebook are prevalent but not polarizing. *Nature* 620.7972, 137–144.

Osmundsen, Mathias, Bor, Alexander, Vahlstrup, Puk Bødker, Bechmann, Anja, and Petersen, Michael Bang (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on Twitter. *American Political Science Review* 115.3, 999–1015.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Pérez, Efrén O. and Tavits, Margit (2022). *Voicing Politics: How Language Shapes Public Opinion*. Princeton, NJ: Princeton University Press.

Reny, Tyler T. and Newman, Benjamin J. (2021). The Opinion-Mobilizing Effect of Social Protest against Police Violence: Evidence from the 2020 George Floyd Protests. *American Political Science Review* 115.4, 1499–1507.

Sears, David O. (1993). Symbolic Politics: A Socio-Psychological Theory.

Sun, Kai, Yao, Yifan, Zhou, Hanwen, Zhang, Jian, Xu, Qinghua, Han, Kaisheng, Li, Shikun, Ma, Yue, Liu, Chunping, Song, Qiaoqiao, et al. (2023). Head-to-tail: How knowledgeable are large language models (LLM)? A.K.A. will LLMs replace knowledge graphs? In: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 4562–4573.

Sun, Wanning and Yu, Haiqing (2023). *Digital Transnationalism: Chinese-Language Media in Australia*. Leiden: Brill.

Theocharis, Yannis, Boulianne, Shelley, Koc-Michalska, Karolina, and Bimber, Bruce (2022). Platform Affordances and Political Participation: How Social Media Reshape Political Engagement. *West European Politics* 46.4, 788–811.

Vaccari, Cristian and Valeriani, Augusto (2021). *Outside the Bubble: Social Media and Political Participation in Western Democracies*. New York: Oxford University Press.

Velez, Yamil and Liu, Patrick (2024). Confronting Core Issues: A Critical Assessment of Attitude Polarization Using Tailored Experiments. *American Political Science Review*. Published online ahead of print, 1–18.

Velez, Yamil, Porter, Ethan, and Wood, Thomas J. (2023). Latino-Targeted Misinformation and the Power of Factual Corrections. *The Journal of Politics* 85.2, 789–794.

Vosoughi, Soroush, Roy, Deb, and Aral, Sinan (2018). The spread of true and false news online. *Science* 359.6380, 1146–1151.

Xu, Keping S. (2021). Repress, co-opt, or nudge? The effect of online information control on citizen trust in government. *Comparative Political Studies* 54.14, 2551–2584.

Xu, Xu, Kostka, Genia, and Cao, Xun (2022). Information Control and Public Support for Social Credit Systems in China. *The Journal of Politics*.

Zaller, John R. (1992). *The Nature and Origins of Mass Opinion*. Cambridge Studies in Public Opinion and Political Psychology. Cambridge University Press.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Appendix

A Python Code for Data Collection and LLM Labeling

Appendix A.1 Implementation Architecture

My automated text classification pipeline leverages OpenAI's GPT-4.1 API to process large-scale social media datasets efficiently. The implementation prioritizes robustness, cost-effectiveness, and data integrity through several key architectural decisions. Table A.1 summarizes the core technical features that ensure reliable classification across diverse content types and platform contexts.

Table A.1: Key Features of LLM Classification Implementation

Feature	Implementation Details
Model Selection	GPT-4.1 via OpenAI API with deterministic settings
Error Handling	Exception handling with graceful degradation
Rate Limit Management	Exponential backoff with automatic retry mechanism
Progress Management	Checkpoint saving with resume functionality

```python
resp = client.chat.completions.create(
    model="gpt-4.1",
    messages=[{"role": "user", "content": build_prompt(content)}],
    temperature=0  # For consistent results
)
```

```python
try:
    parsed = extract_json(response_content)
    result = validate_response(parsed)
except (json.JSONDecodeError, KeyError, ValueError) as e:
    return default_values()
except (APIError, RateLimitError) as e:
    return default_values()
```

```python
except RateLimitError as e:
    retry_count += 1
    time.sleep(backoff)
    backoff = min(backoff * 2, max_backoff)
```

```python
for idx in tqdm(range(start_idx, len(df))):
    result = classify_text(client, content, model)
    df.loc[idx, output_columns] = result.values()

    if (idx + 1) % checkpoint_freq == 0:
        df.to_csv(output_file, index=False)
```

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Appendix A.2 Few-Shot Learning Examples

To ensure consistent application of the coding framework, the prompt includes nine detailed examples that demonstrate key classification principles across different content scenarios. These examples are strategically designed to address common ambiguities in political content analysis, particularly the distinction between policy criticism and personal attacks, the calibration of misinformation severity, and the identification of ingroup/outgroup dynamics.

Table A.2 presents selected examples that illustrate the most challenging aspects of the classification task. Each example includes the original Chinese content, the expected JSON output, and the specific learning objectives it addresses.

Table A.2: Few-Shot Learning Examples

Example Type	Content & Expected Output	Learning Objectives
Personal Attack	**Content:** "(某共和党候选人) 是个笨蛋，他根本不配当总统" **Translation:** "(A Republican candidate) is an idiot, he is completely unfit to be president" **Output:** {"misinfo": 0, "pro_dem": 0, "anti_dem": 0, "pro_rep": 0, "anti_rep": 1, "attack": 1, "ingroup_like": 0, "outgroup_hate": 0}	Distinguish personal insults from policy criticism; Link attacks on party leaders to anti-party sentiment; Recognize absence of group identification
Misinfo	**Content:** "（某民主党候选人）故意关停美国所有的石油管道，这真是对国家犯罪" **Translation:** "(A Democratic Party candidate) deliberately shut down all oil pipelines in America, this is truly a crime against the nation" **Output:** {"misinfo": 0.75, "pro_dem": 0, "anti_dem": 1, "pro_rep": 0, "anti_rep": 0, "attack": 0, "ingroup_like": 0, "outgroup_hate": 0}	Calibrate ordinal misinformation scale; Identify misleading factual claims; Separate strong criticism from personal attacks
Outgroup Hostility	**Content:** "我喜欢共和党但我不喜欢川普，我不喜欢民主党（让人感到恶心）但我支持拜登的政策" **Translation:** "I like the Republican Party but I don't like Trump, I don't like the Democratic Party (it's disgusting) but I support Biden's policies" **Output:** {"misinfo": 0, "pro_dem": 0, "anti_dem": 0, "pro_rep": 0, "anti_rep": 0, "attack": 0, "ingroup_like": 1, "outgroup_hate": 1}	Identify explicit group preferences; Distinguish affective polarization from policy positions; Capture simultaneous ingroup/outgroup attitudes

Appendix A.3 Data Collection Using X's API

My automated data collection pipeline leverages X's Academic Research API v2 to systematically gather historical tweet data with comprehensive error handling and rate limit management. The implementation prioritizes data integrity, efficient resource utilization, and robust checkpoint saving to handle large-scale collection operations across multiple keywords and extended time periods. Table A.3 summarizes the core technical features that ensure reliable data collection across diverse query patterns and API constraints.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

The pipeline processes queries iteratively with intelligent rate limit monitoring and proactive throttling to maximize collection efficiency while respecting API constraints. Key design principles include adaptive delay mechanisms, real-time checkpoint saving, and comprehensive metadata extraction to ensure complete data preservation throughout the collection process.

Table A.3: Key Features of X API Data Collection Implementation

Feature	Implementation Details
API Configuration	**X's Pro Service API v2 with metadata collection**
	`params = { 'query': f"{keyword} lang:zh -is:retweet", 'max_results': 100, 'tweet.fields': 'created_at,author_id,public_metrics...', 'user.fields': 'location,username,verified...', 'expansions': 'author_id' }`
Rate Limit Management	**Proactive monitoring with intelligent throttling**

	`if response.status_code == 429: wait_time = max(1, reset_time - int(time.time()) + 1) time.sleep(wait_time)`
	`# Proactive protection when remaining < 10`
	`if remaining and int(remaining) < 10: time.sleep(900) # Wait 15 minutes to reset`

Error Handling	**Retry mechanism with exponential backoff**
	`@retry(tries=3, delay=1, backoff=2, max_delay=10)` `def search_tweets(self, query, ...): # Robust API call with automatic retries if result is None or result[0] is None: break # Graceful failure handling`
Checkpoint Saving	**Real-time data preservation after each API call**

	`# Save ~100 tweets after each API response`
	`formatted_data = self.extract_tweet_data(response)`
	`self.append_to_file(formatted_data, filename)`
	`mode = 'a' if os.path.exists(filename) else 'w'`
	`df.to_csv(filename, mode=mode, header=header, index=False)`

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

B Inter-Coder Agreement between LLM Coding and Expert Coding

Comprehensive Expert vs LLM Coding Agreement and Accuracy Metrics

		Pearson Correlation		Spearman Correlation		Agreement		Category-Level Performance	
Variable	N	Pearson r	p-value	Spearman ρ	p-value	Cohen's κ	p-value	Overall Accuracy	Avg F1 Score
Misinformation Variables									
Misinformation (Raw)	500	0.814	0	0.789	0	0.717	NA	0.906	0.816
Misinformation (Recoded)	500	0.822	0	0.792	0	0.721	NA	0.906	0.816
Partisan Sentiment									
Republican Stance	500	0.762	0	0.760	0	0.728	NA	0.902	0.828
Democrat Stance	500	0.682	0	0.685	0	0.691	NA	0.940	0.757
Affective Polarization									
Ingroup Liking	500	0.759	0	0.759	0	0.731	NA	0.970	0.865
Outgroup Hatred	500	0.820	0	0.820	0	0.813	NA	0.974	0.906

Figure A.1: Inter-coder Agreement between LLM Coding and Expert Coding

Appendix B presents comprehensive validation analyses comparing LLM-generated labels with expert annotations across six key variables in our content analysis framework. These results demonstrate the reliability and validity of our automated labeling approach while highlighting important patterns in inter-coder agreement.

The validation results reveal consistently strong correlations between expert and LLM coding across all measured dimensions. Pearson correlations range from 0.682 to 0.822, with all coefficients achieving statistical significance at $p < 0.001$. Misinformation coding demonstrates the highest correlation ($r = 0.822$ for recoded categories, $r = 0.814$ for raw scores), followed closely by outgroup hatred detection ($r = 0.820$). Republican stance identification shows robust correlation ($r = 0.762$), while Democratic stance presents a slightly lower but still substantial correlation ($r = 0.682$). Ingroup liking detection achieves strong correlation ($r = 0.759$), indicating reliable automated identification of positive affective content.

Spearman rank correlations closely mirror the Pearson results, ranging from 0.685

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

to 0.820, suggesting that the relationship between expert and LLM coding remains robust across different distributional assumptions. The consistency between Pearson and Spearman correlations indicates that the strong agreement is not driven by outliers or non-linear relationships.

Cohen's kappa coefficients provide additional evidence of substantial inter-coder agreement, with values ranging from 0.691 to 0.813. According to conventional interpretation guidelines, all kappa values exceed 0.60, indicating substantial agreement, with most approaching the 0.80 threshold for near-perfect agreement. Outgroup hatred achieves the highest kappa (0.813), followed by misinformation detection (0.721 for recoded categories), and Republican stance identification (0.728).

Category-level accuracy metrics demonstrate the practical reliability of LLM coding for research applications. Overall accuracy ranges from 0.902 to 0.974 across all variables, with affective polarization variables achieving the highest accuracy rates (0.970 for ingroup liking, 0.974 for outgroup hatred). This superior performance likely reflects the binary nature of these variables, which simplifies the classification task compared to ordinal measures.

Misinformation detection achieves 90.6% overall accuracy with an average F1 score of 0.816, indicating balanced performance across different misinformation categories. Partisan sentiment identification shows strong performance for both Republican stance (90.2% accuracy, F1 = 0.828) and Democratic stance (94.0% accuracy, F1 = 0.757). The higher accuracy for Democratic stance detection, despite lower correlation coefficients, suggests that while the LLM may have different threshold sensitivities, it reliably identifies clear cases of Democratic content.

C Time series trends of information production

Appendix C.1 Study 1

Figure A.2 presents the time-series trends of Chinese users' content creation, disaggregated by misinformation levels. The platforms exhibit markedly different response patterns: X shows a dramatic spike in content creation on the day of the assassination attempt, while Weibo displays only a modest increase. On both platforms, mostly accurate information dominates the ecosystem. Still, users also create information that is either partially misleading or mostly inaccurate. On X, misinformation of each category is greater than its corresponding category on Weibo.

Figure A.3 displays the time-series trends of Chinese users' content creation, disaggregated by partisan orientation. Partisan messaging patterns reveal interesting platform differences. On X, while non-partisan messages dominate the discourse, partisan messages still comprise a substantial proportion of the discussion. Conversely, Weibo content

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

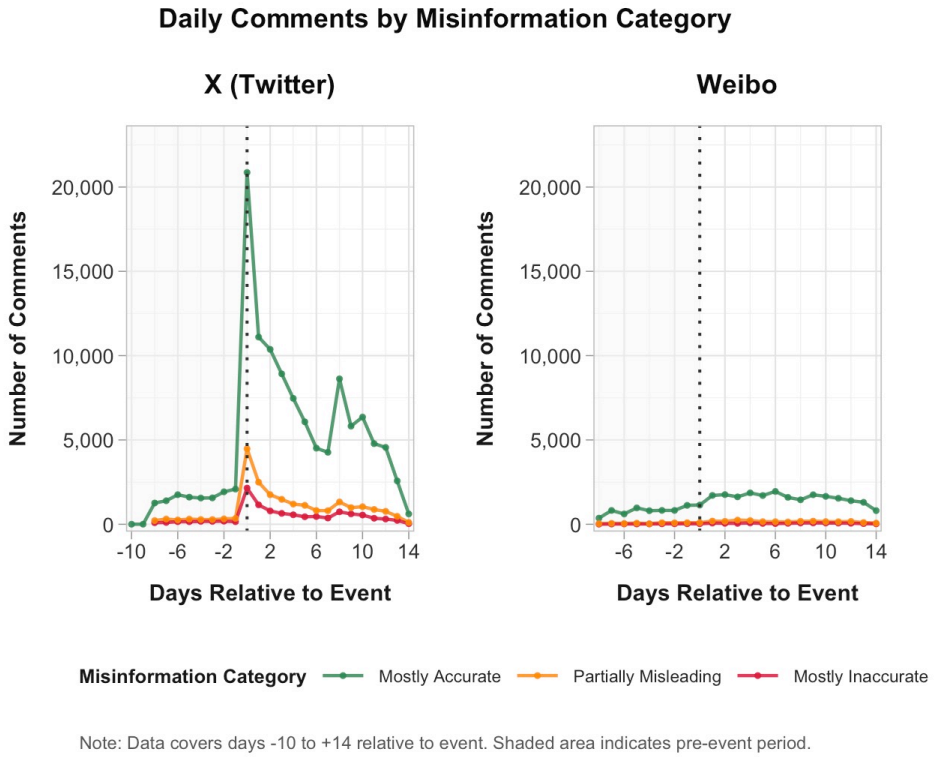


Figure A.2: Time-Series Trends of Post Counts by Misinformation Level

remains predominantly non-partisan, and the proportion of partisan messages is relatively tiny.

Appendix C.2 Biden Withdrawal as Potential Confound

A potential threat to my identification strategy emerges on day 8 (July 21), when President Biden withdrew from the presidential race. This timing coincides with the latter portion of our event study window, raising concerns that renewed political discussion might confound our ability to isolate the assassination attempt's effects. To address this validity concern, I examine discussion patterns on X (Twitter) following Biden's withdrawal, analyzing both absolute volumes and compositional changes across content categories.

Figure A.4 presents absolute discussion volumes by content type starting from Biden's withdrawal date. Trump-related content (blue line) dominates the discussion landscape across both user groups, experiencing an immediate spike coinciding with the withdrawal announcement—reaching approximately 1,600 posts among influencers and 4,800 posts among users on day 8. This surge demonstrates that Biden's exit immediately intensified Trump-focused discourse, as expected given the electoral implications of the withdrawal.

However, assassination-related content (orange line) maintains stable, low-volume patterns throughout the post-withdrawal period, with daily posts consistently remaining

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

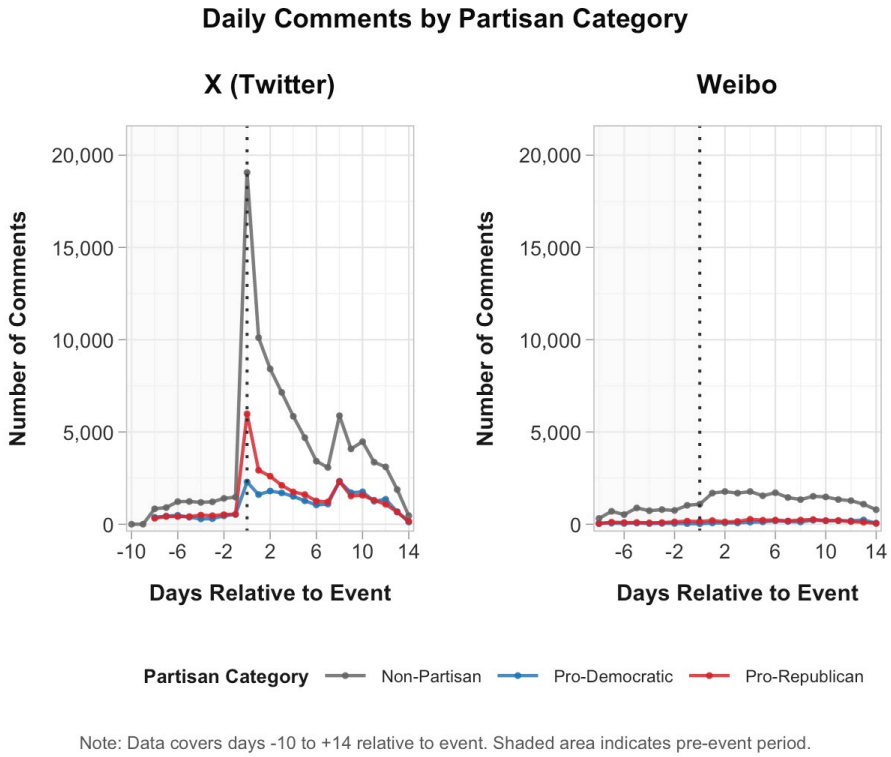


Figure A.3: Time-Series Trends of Post Counts by Partisan Orientation

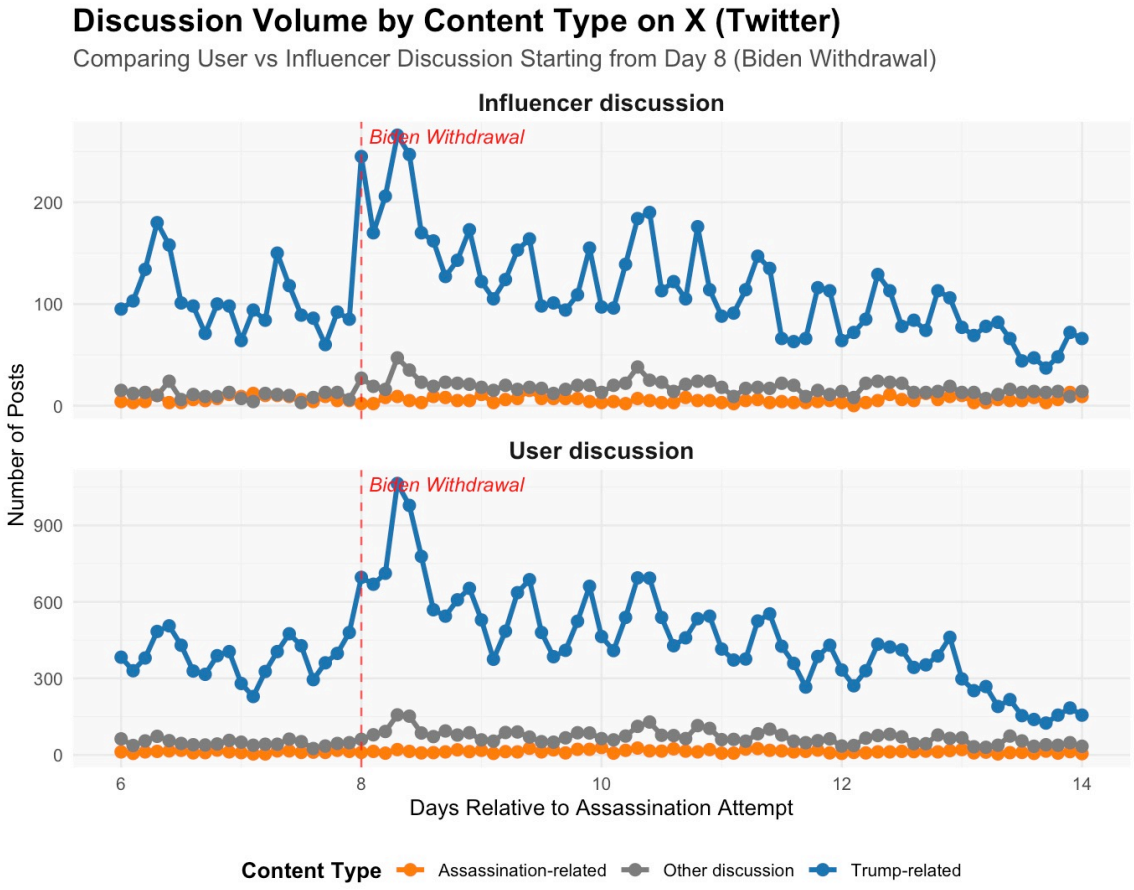


Figure A.4: Discussion Volumes by Content Type Following Biden Withdrawal

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

below 500 across both user groups. This stability indicates that Biden's exit did not trigger renewed assassination-related discussion or conflate the two political events in users' discourse patterns. The persistence of assassination-related content at modest but consistent levels suggests organic continuation of assassination-focused discourse rather than Biden-withdrawal-induced conversation.

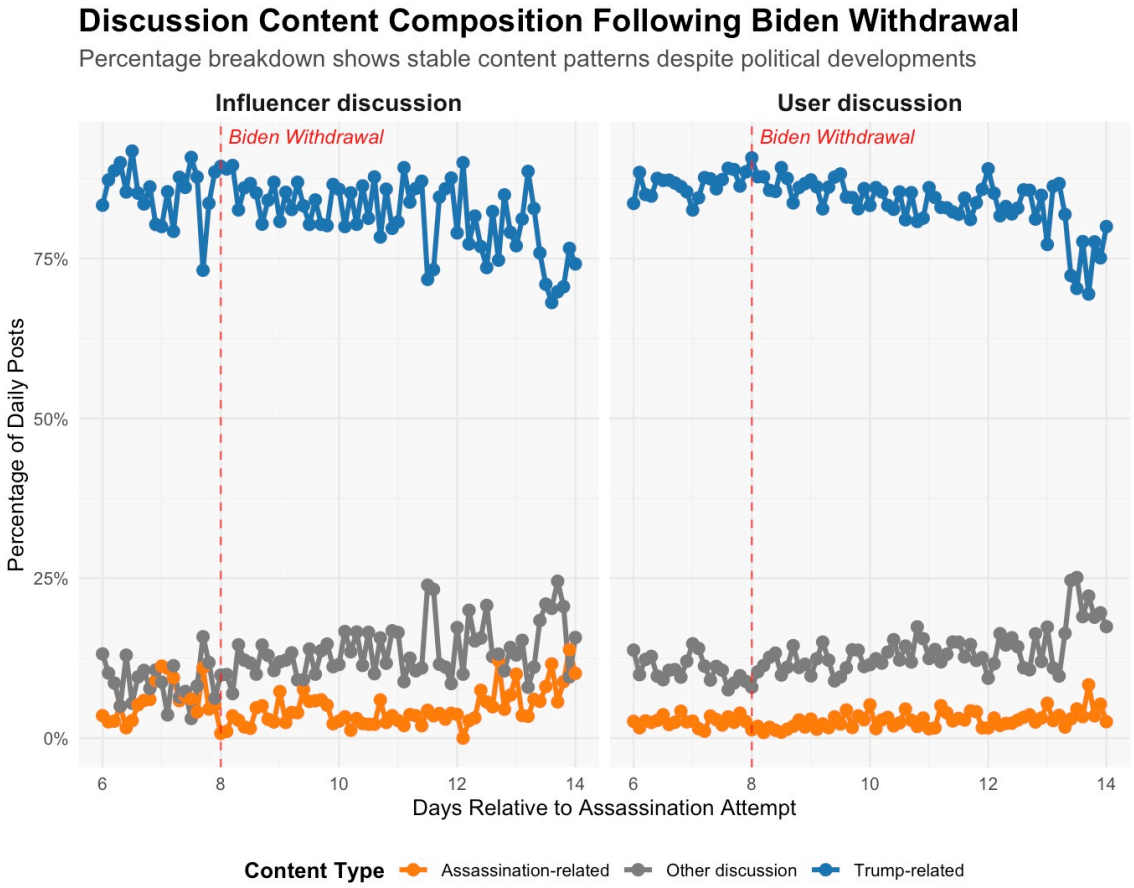


Figure A.5: Discussion Content Composition Following Biden Withdrawal

Figure A.5 provides evidence by examining discussion composition rather than absolute volumes. The proportion-based analysis reveals that Trump-related content (blue line) consistently dominates discourse across both user groups, maintaining approximately 70-80% of daily posts throughout the observation period. Critically, this proportional dominance remains largely stable both before and after Biden's withdrawal, with only modest fluctuations that do not represent fundamental compositional shifts.

Most importantly for my identification strategy, assassination-related content (red line) exhibits remarkable proportional stability, consistently comprising 10-25% of influencer discussions and 5-15% of user discussions across all observed days. The assassination-related discourse shows no systematic increase following Biden's withdrawal, directly contradicting concerns that Biden's exit might artificially inflate assassination-related discussion. The "other discussion" category (green line) similarly maintains steady representation at approximately 10-20% across both user types.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Appendix C.3 Study 2

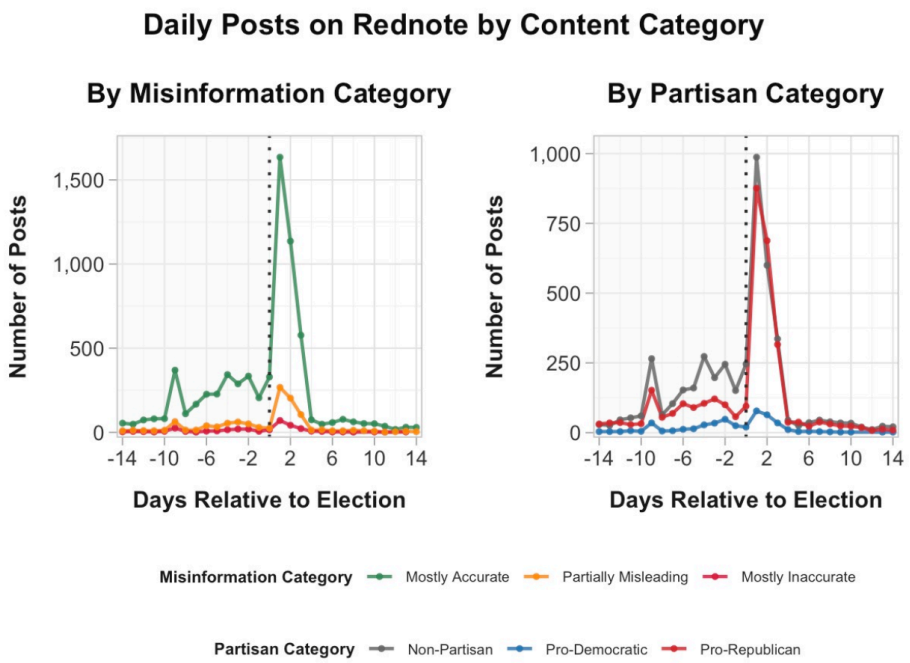


Figure A.6: Daily Posts on Rednote by Content Category. The figure shows posting activity across misinformation and partisan categories relative to the 2024 US Election.

Figure A.6 displays posting activity on Rednote across misinformation and partisan content categories in the days surrounding the 2024 US Election. The left panel shows posts categorized by misinformation accuracy (mostly accurate in green, partially misleading in orange, mostly inaccurate in red), while the right panel categorizes posts by partisan lean (non-partisan in gray, pro-Democratic in blue, pro-Republican in red). Both panels reveal a sharp spike in posting activity around election day (day 0), with mostly accurate and non-partisan content dominating the volume.

Figure A.7 shows comment activity on Rednote using the same categorization scheme as Figure A.5. Comment patterns display more pronounced engagement spikes than posting activity, with two notable peaks occurring approximately 10 days before the election and 2 days after. The volume of comments on mostly accurate and non-partisan content significantly exceeds other categories, particularly during peak engagement periods around the election.

D Full Models

Appendix D.1 Full Model for Figure 3

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

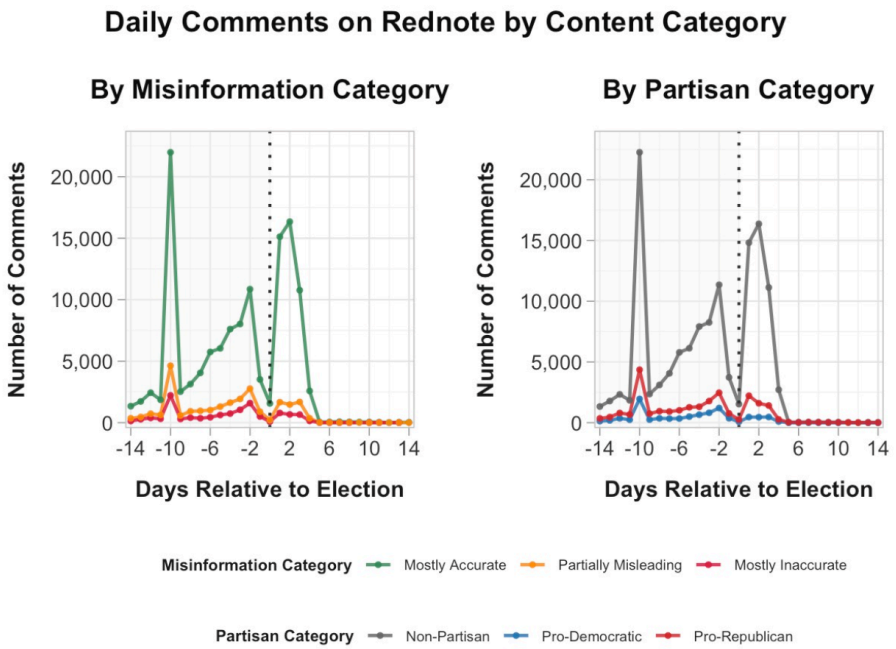


Figure A.7: Daily Comments on Rednote by Content Category. Comment activity displayed more pronounced spikes than posting activity, with two notable peaks occurring around days -10 and +2 relative to the election.

Table A.4: Content Production Analysis: Complete Model Results

	Misinformation		Partisanship	
	X	Weibo	X	Weibo
Main Effects				
Intercept	6.899***	1.862***	5.467***	1.782***
	(0.929)	(0.078)	(0.711)	(0.075)
Time	9.563***	0.602***	8.737***	0.613***
	(1.196)	(0.071)	(1.093)	(0.073)
Partially Misleading	-3.376***	-0.676***	–	–
	(0.801)	(0.074)	–	–
Mostly Inaccurate	-2.585*	-0.648***	–	–
	(1.199)	(0.114)	–	–
Pro-Republican	–	–	-0.365	-0.462***
	–	–	(1.335)	(0.078)
Pro-Democratic	–	–	0.922	-0.560***
	–	–	(1.114)	(0.098)
Netizen	-4.233***	-0.687***	-3.408***	-0.610***
	(0.954)	(0.079)	(0.723)	(0.077)
IP: Undisclosed	-3.868***	0.896*	-2.911***	0.811*

Continued on next page

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Table A.4 continued from previous page

	Misinformation		Partisanship	
	X	Weibo	X	Weibo
	(0.972)	(0.376)	(0.732)	(0.351)
IP: China	-4.577***	–	-3.308***	–
	(0.984)	–	(0.782)	–
IP: International	-3.524***	1.325***	-2.568***	1.182***
	(1.039)	(0.499)	(0.794)	(0.446)
Two-way Interactions				
Time × Partially Misleading	-6.021***	-0.376***	–	–
	(0.999)	(0.077)	–	–
Time × Mostly Inaccurate	-8.391***	-0.520***	–	–
	(1.327)	(0.113)	–	–
Time × Pro-Republican	–	–	-5.089***	-0.469***
	–	–	(1.055)	(0.086)
Time × Pro-Democratic	–	–	-5.741***	-0.486***
	–	–	(1.463)	(0.106)
Time × Netizen	-7.928***	-0.554***	-7.209***	-0.563***
	(1.207)	(0.073)	(1.102)	(0.075)
Time × IP: Undisclosed	-6.895***	-0.740***	-5.947***	-0.608**
	(1.219)	(0.232)	(1.112)	(0.242)
Time × IP: China	-7.650***	–	-6.827***	–
	(1.242)	–	(1.133)	–
Time × IP: International	-5.724***	-0.416	-4.719***	-0.392
	(1.283)	(0.376)	(1.191)	(0.341)
Partially Misleading × Netizen	2.483**	0.580***	–	–
	(0.819)	(0.080)	–	–
Mostly Inaccurate × Netizen	1.503	0.496***	–	–
	(1.221)	(0.117)	–	–
Pro-Republican × Netizen	–	–	0.179	0.378***
	–	–	(1.345)	(0.084)
Pro-Democratic × Netizen	–	–	0.024	0.425***
	–	–	(1.188)	(0.100)
Partially Misleading × IP: Undisclosed	2.490**	-0.332	–	–
	(0.846)	(0.405)	–	–
Mostly Inaccurate × IP: Undisclosed	1.599	-0.610	–	–
	(1.253)	(0.564)	–	–

Continued on next page

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Table A.4 continued from previous page

	Misinformation		Partisanship	
	X	Weibo	X	Weibo
Pro-Republican × IP: Undisclosed	–	–	0.461	-0.239
	–	–	(1.375)	(0.317)
Pro-Democratic × IP: Undisclosed	–	–	-0.105	-0.324
	–	–	(1.337)	(0.341)
Partially Misleading × IP: China	2.554**	–	–	–
	(0.900)	–	–	–
Mostly Inaccurate × IP: China	1.887	–	–	–
	(1.303)	–	–	–
Pro-Republican × IP: China	–	–	0.374	–
	–	–	(1.581)	–
Pro-Democratic × IP: China	–	–	-1.224	–
	–	–	(1.200)	–
Partially Misleading × IP: International	1.410	-1.324**	–	–
	(0.913)	(0.490)	–	–
Mostly Inaccurate × IP: International	1.286	-1.540***	–	–
	(1.422)	(0.506)	–	–
Pro-Republican × IP: International	–	–	-0.778	-0.675
	–	–	(1.385)	(0.391)
Pro-Democratic × IP: International	–	–	0.317	-0.849
	–	–	(1.539)	(0.455)
Netizen × IP: Undisclosed	3.172**	-0.903*	2.525**	-0.832*
	(0.998)	(0.378)	(0.744)	(0.352)
Netizen × IP: China	3.265***	–	2.564**	–
	(1.010)	–	(0.796)	–
Netizen × IP: International	2.490**	-1.425***	2.045**	-1.294***
	(1.066)	(0.500)	(0.809)	(0.447)
Three-way Interactions				
Time × Partially Misleading × Netizen	5.263***	0.343***	–	–
	(1.017)	(0.082)	–	–
Time × Mostly Inaccurate × Netizen	7.478***	0.556***	–	–
	(1.347)	(0.119)	–	–
Time × Pro-Republican × Netizen	–	–	4.739***	0.415***
	–	–	(1.068)	(0.092)
Time × Pro-Democratic × Netizen	–	–	4.984***	0.463***

Continued on next page

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Table A.4 continued from previous page

	Misinformation		Partisanship	
	X	Weibo	X	Weibo
	–	–	(1.493)	(0.109)
Time × IP: Undisclosed × Netizen	6.359***	0.726***	5.405***	0.595**
	(1.231)	(0.234)	(1.122)	(0.244)
Time × IP: China × Netizen	6.755***	–	5.902***	–
	(1.255)	–	(1.145)	–
Time × IP: International × Netizen	5.462***	0.408	4.286***	0.378
	(1.297)	(0.377)	(1.203)	(0.342)
Observations	34,892	18,758	34,892	18,758
R^2	0.046	0.061	0.042	0.060
Adjusted R^2	0.045	0.059	0.042	0.059

Notes: Robust standard errors clustered at username level in parentheses. Dummy coding used for categorical variables. Misinformation models: Reference categories are Mostly Accurate (content), Influencer (type), US (IP location). Partisanship models: Reference categories are Non-partisan (content), Influencer (type), US (IP location for X), China domestic (IP location for Weibo). **Excluded coefficients**: Four-way interactions and some higher-order terms are omitted for space. Full results available upon request. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The intercept coefficients represent baseline comment counts for US-based influencers during the pre-event period. X shows substantially higher baseline activity (6.899 for misinformation, 5.467 for partisanship) compared to Weibo (1.862 and 1.782 respectively). Time coefficients demonstrate dramatic post-event increases, with X users showing larger increases (9.563 and 8.737) than Weibo users (0.602 and 0.613).

Content main effects reveal platform differences in misinformation production. On X, both partially misleading (-3.376) and mostly inaccurate (-2.585) content show negative coefficients relative to accurate content. Weibo shows similar patterns with stronger effects (-0.676 and -0.648). For partisanship, X shows non-significant effects while Weibo shows significant negative coefficients for both partisan categories.

Netizen coefficients are consistently negative across all models, indicating lower baseline activity than influencers. X shows larger effects (-4.233 for misinformation, -3.408 for partisanship) than Weibo (-0.687 and -0.610). Geographic patterns differ by platform: X shows negative coefficients for all non-US locations, while Weibo shows positive coefficients for international and undisclosed users.

Time interactions reveal differential content responses to the temporal shock. Both platforms show negative time-content interactions, indicating smaller post-event surges

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

for problematic content. Time-netizen interactions are large and negative on both platforms, with X showing stronger effects. Time-geography interactions on X show the largest negative effects for China users, followed by undisclosed and international users.

Three-way interactions capture complex dynamics between time, content, and user characteristics. Time-content-netizen interactions on X range from 4.739 to 7.478, indicating netizens show particularly strong increases in specific content types post-event. Time-geography-netizen interactions show non-US netizens experience large post-event increases, with coefficients ranging from 4.286 to 6.755 on X.

Appendix D.2 Two Period Diff-in-Diff Estimates by IP Locations

In the main text, I present two period DiD estimates in the full sample, as well as event-study plots for the full sample and by different IP subgrouops. Here, I also present two period DiD estimates by IP subgroups as an additional robustness check.

Table A.5: Difference-in-Differences Analysis: Platform Effects by IP Location

	Outcome Variables					
	Post Count	Affective Polarization	Misinformation Posts	Republican Content	Democratic Content	Attack Language
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline (Weibo Pre-Event)	1.187***	0.066***	0.086***	0.125***	0.077***	0.109***
	(0.014)	(0.005)	(0.005)	(0.007)	(0.005)	(0.006)
Post-Event Effect (Weibo)	0.053***	0.035***	0.057***	0.018**	0.048***	0.022***
	(0.013)	(0.006)	(0.006)	(0.008)	(0.006)	(0.006)
Pre-Event Platform Differences:						
US vs Weibo	1.795***	0.607***	0.517***	0.726***	0.917***	0.738***
	(0.244)	(0.074)	(0.066)	(0.083)	(0.171)	(0.093)
China users on X vs Weibo	0.247***	0.213***	0.186***	0.235***	0.185***	0.224***
	(0.073)	(0.046)	(0.040)	(0.069)	(0.033)	(0.033)
International users on X vs Weibo	0.655***	0.283***	0.318***	0.395***	0.262***	0.334***
	(0.110)	(0.034)	(0.047)	(0.067)	(0.032)	(0.038)
Undisclosed IP users on X vs Weibo	0.987***	0.359***	0.377***	0.498***	0.430***	0.488***
	(0.075)	(0.022)	(0.022)	(0.030)	(0.043)	(0.025)
Platform Effects (DiD Estimates):						
Post Event × US Users on X	2.043***	0.493***	0.474***	0.640***	0.130	0.400***
	(0.182)	(0.087)	(0.066)	(0.079)	(0.098)	(0.078)
Post Event × China Users on X	0.820***	-0.001	0.092**	0.130*	0.043	0.057
	(0.096)	(0.049)	(0.045)	(0.073)	(0.037)	(0.037)
Post Event × International Users on X	1.480***	0.190***	0.190***	0.323***	0.183***	0.207***
	(0.121)	(0.042)	(0.044)	(0.066)	(0.045)	(0.050)
Post Event × Undisclosed IP Users on X	1.283***	0.170***	0.237***	0.257***	0.126***	0.163***
	(0.066)	(0.021)	(0.022)	(0.027)	(0.035)	(0.023)
Observations	50,621	50,621	50,621	50,621	50,621	50,621
R-squared	0.033	0.022	0.021	0.026	0.010	0.024

Notes: Robust standard errors clustered by username in parentheses. The baseline represents the average outcome for Weibo users in the pre-event period. Post-Event Effect shows the change in Weibo users' behavior following the focal event. Pre-Event Platform Differences show baseline differences between X (Twitter) IP groups and Weibo users before the event. DiD estimates show the additional change experienced by each X (Twitter) IP group relative to Weibo users following the focal event, net of pre-existing differences. Affective polarization measures the sum of ingroup favoritism and outgroup hostility. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

The baseline results show that Weibo users in the pre-event period averaged 1.187 posts per day. Their baseline engagement with political content was substantial, with coefficients of 0.066 for affective polarization, 0.086 for misinformation posts, 0.125 for Republican content, 0.077 for Democratic content, and 0.109 for attack language. Following the focal event, Weibo users significantly increased their activity levels, posting an additional 0.053 posts on average. Most notably, they increased their engagement across all content types: affective polarization by 0.035, misinformation content by 0.057, Republican content by 0.018, Democratic content by 0.048, and attack language by 0.022.

The pre-event platform differences reveal distinct baseline patterns across IP location groups relative to Weibo users. US-based users showed the strongest differential engagement, posting 1.795 more posts on average and exhibiting substantially higher baseline rates across all content types: affective polarization (0.607), misinformation (0.517), Republican content (0.726), Democratic content (0.917), and attack language (0.738) compared to Weibo users. China-based users showed modest but significant differences in posting frequency (0.247) and demonstrated elevated engagement with affective polarization (0.213), misinformation (0.186), Republican content (0.235), Democratic content (0.185), and attack language (0.224). International users fell between these extremes with moderate increases in posting (0.655) and content engagement across all categories: affective polarization (0.283), misinformation (0.318), Republican content (0.395), Democratic content (0.262), and attack language (0.334). Users with undisclosed IP locations showed strong baseline differences with 0.987 more posts and elevated engagement across all content types: affective polarization (0.359), misinformation (0.377), Republican content (0.498), Democratic content (0.430), and attack language (0.488).

The difference-in-differences estimates reveal dramatically amplified responses to the focal event among X (Twitter) users compared to Weibo users. US-based users exhibited the largest treatment effects, increasing their posting by an additional 2.043 posts beyond the Weibo baseline response. Their engagement with political content intensified substantially, with increases of 0.493 for affective polarization, 0.474 for misinformation posts, and 0.640 for Republican content. Notably, their Democratic content response (0.130) was not statistically significant, while attack language increased significantly by 0.400. China-based users showed more modest amplification effects with additional increases of 0.820 posts, but their responses in content engagement were more limited—only misinformation (0.092) and Republican content (0.130) showed significant increases, while affective polarization showed no effect (-0.001), and Democratic content and attack language showed non-significant changes. International users demonstrated moderate but significant treatment effects across most measures, with increases of 1.480 posts, 0.190 for both affective polarization and misinformation, 0.323 for Republican content, 0.183 for Democratic content, and 0.207 for attack language. Users with undisclosed IP locations showed strong responses with 1.283 additional posts and significant increases across all

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

content categories: 0.170 for affective polarization, 0.237 for misinformation, 0.257 for Republican content, 0.126 for Democratic content, and 0.163 for attack language.

Appendix D.3 Full Model for Figure 6

Table A.6: Regression Results: Post-Election Effects by User Location

	Dependent Variables			
	Note Count	Avg Misinfo	Avg Pro-Rep	Avg Pro-Dem
Main Effects				
Post-Election	−0.305***	0.005	0.178***	−0.036***
	(0.040)	(0.007)	(0.015)	(0.006)
US Location	0.144	0.009	0.051*	0.009
	(0.130)	(0.011)	(0.023)	(0.012)
International Location	0.001	0.008	0.011	0.001
	(0.108)	(0.015)	(0.031)	(0.016)
Interaction Effects				
Post-Election × US	−0.110	0.029	0.002	0.016
	(0.122)	(0.016)	(0.035)	(0.015)
Post-Election × International	−0.002	0.015	0.061	0.007
	(0.106)	(0.020)	(0.043)	(0.019)
Constant	1.378***	0.146***	0.265***	0.053***
	(0.040)	(0.005)	(0.011)	(0.005)
Observations	4,975	4,975	4,975	4,975
R-squared	0.017	0.003	0.038	0.010
Adjusted R-squared	0.016	0.002	0.037	0.009
F-statistic	15.55	3.05	43.16	10.55

Notes: Robust standard errors clustered by user ID in parentheses. Reference category is non-US, non-international location (domestic non-US). * p<0.05, ** p<0.01, *** p<0.001.

The note-level analysis reveals significant post-election shifts in user behavior and content characteristics. Following the election, there was a substantial 0.305-unit decrease in note count per user, indicating reduced overall posting activity. Simultaneously, pro-Republican sentiment increased significantly by 0.178 units while pro-Democrat sentiment decreased by 0.036 units, suggesting a rightward shift in political expression. US-based users demonstrated consistently higher pro-Republican sentiment (0.051 units above the reference group) compared to other locations. Notably, misinformation levels showed no significant main effect changes post-election, though interaction terms suggest potential differential effects by location that warrant further investigation. The interaction between post-election period and international location shows a notable 0.061-unit increase in pro-Republican sentiment, indicating that international users may have responded differently to the election outcome than domestic users.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Table A.7: Comment-Level Analysis: Post-Election Effects by User Location

	Dependent Variables					
	Comment Count	Avg Misinfo	Avg Pro-Rep	Avg Pro-Dem	Avg Funny	Avg Attack
Main Effects						
Post-Election	−0.079***	−0.074***	−0.044***	−0.019***	0.025***	−0.035***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)
US Location	0.086***	0.046***	0.028***	0.007***	−0.047***	0.069***
	(0.008)	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)
International Location	−0.043***	−0.015***	−0.018***	−0.001	0.004	0.018***
	(0.004)	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)
Interaction Effects						
Post-Election × US	−0.086***	0.017***	−0.000	0.006*	−0.022**	−0.027***
	(0.008)	(0.004)	(0.004)	(0.002)	(0.007)	(0.004)
Post-Election × International	0.043***	0.037***	0.029***	0.003	−0.005	−0.006
	(0.004)	(0.004)	(0.004)	(0.002)	(0.009)	(0.004)
Constant	1.079***	0.167***	0.088***	0.030***	0.362***	0.084***
	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Observations	179,800	179,800	179,800	179,800	179,800	179,800
R-squared	0.005	0.022	0.008	0.003	0.002	0.013
Adjusted R-squared	0.005	0.022	0.008	0.003	0.002	0.013
F-statistic	–	878.3	302.0	181.6	91.4	460.2

Notes: Robust standard errors clustered by user ID in parentheses. Reference category is non-US, non-international location (domestic non-US). * p<0.05, ** p<0.01, *** p<0.001. F-statistic not available for Comment Count model.

The comment-level analysis, with substantially more observations (179,800), provides a more granular view of post-election behavioral changes and reveals several striking patterns. Post-election effects show decreased comment volume (-0.079), reduced misinformation (-0.074), decreased pro-Republican sentiment (-0.044), decreased pro-Democrat sentiment (-0.019), increased humorous content (0.025), and reduced attack language (-0.035), suggesting a general shift toward less polarized and more lighthearted discourse. Geographic differences are pronounced: US users exhibit significantly higher levels of misinformation (0.046), pro-Republican sentiment (0.028), pro-Democrat sentiment (0.007), and attack language (0.069), while showing less humorous content (-0.047) compared to the reference group. International users display lower misinformation (-0.015) and pro-Republican sentiment (-0.018) but higher attack language (0.018). The interaction effects reveal important nuances: US users show a significant increase in misinformation (0.017) and pro-Democrat sentiment (0.006) post-election, while international users demonstrate increases in misinformation (0.037) and pro-Republican sentiment (0.029), suggesting differential reactions to the election outcome across geographic locations.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Appendix D.4 Full Model for Figure 7

Table A.8 demonstrates how note-level content characteristics systematically influence comment-level behaviors, with notable temporal shifts following the 2024 US election. The results reveal strong positive associations between note and comment characteristics: notes with medium and high misinformation levels increase comment misinformation probability by 4.8 and 5.7 percentage points respectively, while pro-Republican notes increase Trump support comments by 4.4 percentage points, pro-Democrat notes increase Harris support comments by 1.6 percentage points, and anti-Republican notes increase attack language in comments by 3.9 percentage points.

Post-election effects show consistent decreases across all comment types, with the largest reductions in misinformation (-6.6 percentage points) and attack language (-4.1 percentage points), suggesting a general dampening of inflammatory discourse. Geographic patterns are pronounced: US users exhibit significantly higher rates of misinformation (4.9 percentage points) and attack language (6.3 percentage points) in comments, while showing reduced Harris support (-0.5 percentage points), indicating distinct behavioral patterns by location. The interaction effects reveal important temporal dynamics: the influence of note misinformation on comment misinformation weakened post-election (medium: -2.9 percentage points, high: -2.5 percentage points), and pro-Republican notes had diminished effects on Trump support comments (-3.1 percentage points), suggesting that the election outcome may have reduced the amplification of certain content types through the note-comment pathway.

On the other hand, Table A.9 reveals substantial heterogeneity in how note characteristics influence comment behavior across different IP locations, demonstrating that the note-to-comment transmission mechanisms vary significantly by user geography. The main effects show that note characteristics consistently predict corresponding comment behaviors: medium and high misinformation notes increase comment misinformation by 4.0 and 2.5 percentage points respectively, pro-Republican notes increase Trump support comments by 3.2 percentage points, pro-Democrat notes increase Harris support by 2.3 percentage points, and anti-Republican notes increase attack language by 5.5 percentage points. However, the interaction effects reveal striking geographic variations in these relationships.

Most notably, high misinformation content shows dramatically amplified effects for international users, with an additional 8.7 percentage point increase in comment misinformation beyond the baseline effect, while US users show a more moderate amplification of 4.6 percentage points. This suggests that international users are particularly susceptible to misinformation transmission through the note-comment pathway.

Conversely, pro-Democratic content shows reduced effectiveness outside of China, with both US users (-1.0 percentage points) and international users (-1.1 percentage

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

points) showing significantly weaker Harris support responses to pro-Democrat notes. Similarly, anti-Republican content shows diminished attack language effects for both US (-2.7 percentage points) and international (-2.2 percentage points) users compared to Chinese users. Geographic baseline differences are substantial: US users exhibit higher rates of misinformation (4.5 percentage points), Trump support (1.2 percentage points), and attack language (6.5 percentage points) in their comments, while international users show lower misinformation (-0.6 percentage points) and Trump support (-1.9 percentage points) but higher attack language (2.0 percentage points).

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Table A.8: Note Characteristics Predicting Comment Behavior: Pre vs Post-Election Effects

	Comment-Level Dependent Variables			
	Comment Misinfo	Comment Trump Support	Comment Harris Support	Comment Attack
Main Effects				
Post-Election	−0.066***	−0.038***	−0.015***	−0.041***
	(0.002)	(0.002)	(0.001)	(0.001)
Note Characteristics				
Note Misinfo: Medium	0.048***	–	–	–
	(0.003)			
Note Misinfo: High	0.057***	–	–	–
	(0.004)			
Note Pro-Republican	–	0.044***	–	–
		(0.002)		
Note Pro-Democrat	–	–	0.016***	–
			(0.002)	
Note Anti-Republican	–	–	–	0.039***
				(0.004)
Location Controls				
US Location	0.049***	0.013***	−0.005***	0.063***
	(0.002)	(0.002)	(0.001)	(0.002)
International Location	−0.002	−0.016***	−0.007***	0.019***
	(0.002)	(0.002)	(0.001)	(0.002)
Undisclosed Location	−0.038***	−0.048***	−0.020***	−0.005
	(0.003)	(0.003)	(0.001)	(0.003)
Post-Election Interactions				
Post-Election × Note Medium Misinfo	−0.029***	–	–	–
	(0.004)			
Post-Election × Note High Misinfo	−0.025**	–	–	–
	(0.007)			
Post-Election × Note Pro-Republican	–	−0.031***	–	–
		(0.003)		
Post-Election × Note Pro-Democrat	–	–	0.001	–
			(0.006)	
Post-Election × Note Anti-Republican	–	–	–	−0.010
				(0.009)
Constant	0.156***	0.081***	0.022***	0.084***
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	171,573	171,573	171,573	171,573
R-squared	0.028	0.012	0.005	0.015
Adjusted R-squared	0.028	0.012	0.004	0.015

Notes: Robust standard errors clustered by user ID in parentheses. Reference categories: Pre-election period, Note misinformation low level, China location. All models control for IP location. * p<0.05, ** p<0.01, *** p<0.001.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Table A.9: IP Location Heterogeneity in Note-Comment Relationships

	Comment-Level Dependent Variables			
	Comment Misinfo	Comment Trump Support	Comment Harris Support	Comment Attack
Main Effects				
Note Misinfo: Medium	0.040***	–	–	–
	(0.003)			
Note Misinfo: High	0.025***	–	–	–
	(0.005)			
Note Pro-Republican	–	0.032***	–	–
		(0.002)		
Note Pro-Democrat	–	–	0.023***	–
			(0.003)	
Note Anti-Republican	–	–	–	0.055***
				(0.006)
Location Effects				
US Location	0.045***	0.012***	−0.004***	0.065***
	(0.002)	(0.002)	(0.001)	(0.002)
International Location	−0.006**	−0.019***	−0.006***	0.020***
	(0.002)	(0.002)	(0.001)	(0.002)
Post-Election	−0.072***	−0.048***	−0.015***	−0.041***
	(0.001)	(0.001)	(0.001)	(0.001)
IP Location Interactions				
Note Medium Misinfo × US	0.010*	–	–	–
	(0.005)			
Note High Misinfo × US	0.046***	–	–	–
	(0.008)			
Note Medium Misinfo × International	0.001	–	–	–
	(0.005)			
Note High Misinfo × International	0.087***	–	–	–
	(0.011)			
Note Pro-Republican × US	–	0.004	–	–
		(0.004)		
Note Pro-Republican × International	–	0.010*	–	–
		(0.004)		
Note Pro-Democrat × US	–	–	−0.010*	–
			(0.005)	
Note Pro-Democrat × International	–	–	−0.011**	–
			(0.004)	
Note Anti-Republican × US	–	–	–	−0.027**
				(0.009)
Note Anti-Republican × International	–	–	–	−0.022*
				(0.010)
Constant	0.159***	0.084***	0.022***	0.083***
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	162,591	162,591	162,591	162,591
R-squared	0.029	0.010	0.004	0.015
Adjusted R-squared	0.029	0.010	0.004	0.015

Notes: Robust standard errors clustered by user ID in parentheses. Reference categories: China location, Note misinformation low level, Pre-election period. Undisclosed IP locations excluded from analysis. * p<0.05, ** p<0.01, *** p<0.001.

https://doi.org/10.33774/apsa-2025-g9w01 **ORCID:** https://orcid.org/0009-0007-3491-3480 Content not peer-reviewed by APSA. **License:** All Rights Reserved