



# Observation of Information Manipulations against TikTok Ban

## Table of Contents

<b>Table of Contents</b> .....	<b>1</b>
<b>Glossary</b> .....	<b>4</b>
<b>Executive Summary</b> .....	<b>6</b>
<b>Introduction</b> .....	<b>7</b>
<b>Methodology</b> .....	<b>7</b>
<b>Building Similarity Nodes Between User Accounts</b> .....	<b>7</b>
<b>User Feature Construction</b> .....	<b>8</b>
<b>User Behaviour Features</b> .....	<b>8</b>
<b>Co-occurrence Features</b> .....	<b>8</b>
<b>User Clustering</b> .....	<b>8</b>
<b>User Similarity Evaluation</b> .....	<b>8</b>
<b>User Clustering</b> .....	<b>9</b>
<b>Group Analysis</b> .....	<b>9</b>
<b>Opinion Clustering</b> .....	<b>9</b>
<b>Stance Detection and Narrative Summary</b> .....	<b>9</b>
<b>Data Coverage</b> .....	<b>9</b>
<b>Major Timeline and Most Representable Battlefields</b> .....	<b>10</b>
<b>2024, January: Restrictions Imposed by States on Minors</b> .....	<b>10</b>
<b>2024, February: The Biden Campaign Joins TikTok</b> .....	<b>12</b>
<b>2024, March: The House of Representatives Passed H.R. 7521</b> .....	<b>13</b>
Narratives on the legislative process surrounding H.R. 7521's passage by troll users by platforms..	16
Twitter Narratives.....	16
Youtube Narratives.....	17
Weibo Narratives.....	17
TikTok Narratives.....	18
<b>Trolls Echo PRC State-affiliated Media</b> .....	<b>19</b>
<b>Main Troll Groups</b> .....	<b>20</b>
<b>Troll Group: Twitter#2774</b> .....	<b>20</b>
<b>Abnormal Behaviors</b> .....	<b>20</b>
<b>Operated Stories</b> .....	<b>20</b>
<b>Targets of Troll Activities</b> .....	<b>22</b>
<b>Troll Group: Twitter#6257</b> .....	<b>22</b>
<b>Abnormal Behaviors</b> .....	<b>22</b>
<b>Operated Stories</b> .....	<b>23</b>
<b>Targets of Troll Activities</b> .....	<b>24</b>
<b>Troll Group: YouTube #253</b> .....	<b>24</b>
<b>Abnormal Behaviors</b> .....	<b>24</b>
<b>Operated Stories</b> .....	<b>25</b>
<b>Targets of Troll Activities</b> .....	<b>26</b>
<b>Operational Examples of Troll Groups</b> .....	<b>26</b>
<b>DISARM Techniques Used by Troll Groups</b> .....	<b>27</b>
<b>The Infodemic Platform</b> .....	<b>28</b>



## Glossary

Term	Explanation
<b>Troll Account</b>	Taiwan AI Labs employs large language models to analyze accounts on social media platforms, identifying accounts that frequently comment on the same posts together, indicating coordinated behavior. These accounts exhibit long-term similarities in their commenting patterns, suggesting they are not controlled by natural persons but are likely automated or manipulated, thus termed “Troll Accounts.”
<b>Troll Group</b>	When Troll Accounts show long-term similarities in commenting patterns and signals, they are grouped into a “Troll Group.” These groups can be analyzed for the events they participate in and the targets they manipulate, providing insights into the political forces they may serve.
<b>Event</b>	When an event occurs, it generates extensive news coverage and social media discussions, including posts and videos. Taiwan AI Labs uses large language models to organize these reactions into an “Event,” facilitating the observation of social media manipulation related to the event.
<b>Story</b>	Events can develop over time, linking many related events into a continuous narrative. Through classification with large language models, these interconnected events can be organized into a “Story,” summarizing the coordinated manipulation and related news across a prolonged period for each story, allowing for the observation of long-term collaborative operations.
<b>Media Volume</b>	Media Volume refers to the amount of media presence, calculated by the number of news reports.
<b>(PRC) State-affiliated Media</b>	(PRC) State-affiliated Media denotes media outlets whose content is controlled or censored by the government of the People's Republic of China.
<b>Community Volume</b>	Community Volume represents the volume on social media platforms, encompassing the total number of comments observed from both troll accounts and regular accounts.
<b>Troll Volume</b>	Troll Volume pertains to the volume of comments made by troll accounts.
<b>User Behavior Features</b>	Analysis of social media data reveals a series of columns that represent user behavior features, such as the ‘destination of user interactions’ (post_id or video_id), the ‘time of user actions’, and the ‘domain of shared links by users’, among others. These data are subsequently utilized for user clustering.
<b>Co-occurrence Features</b>	Co-occurrence features aim to identify users who frequently engage with the same topics or respond to the same articles, appearing together in the same context to create a fabricated volume, a common characteristic of troll accounts. Through this method, we can identify troll accounts and cluster them into troll groups.
<b>User Clustering</b>	Taiwan AI Labs analyzes the relationship between pairs of accounts based on a series of signals and assigns a score. If the score exceeds a certain threshold, a connection is established. If multiple accounts are connected, they are clustered into a troll group.

<b>Group Analysis</b>	Taiwan AI Labs uses Taiwan LLM, a large language model pre-trained in Taiwanese dialects, to classify the comments and opinions of troll groups, identifying their main narratives and analyzing the primary information manipulated by troll groups and their underlying intentions.
<b>Topic Engagement</b>	Taiwan AI Labs employs large language models to analyze community platform posts and comments related to news, identifying traces of message manipulation by troll groups. This clarifies which topics troll groups participate in and manipulate discussions on.
<b>Operation Methods</b>	Taiwan AI Labs utilizes the DISARM Framework to analyze the methods and intentions behind the information operations conducted by troll groups.
<b>Leverage Existing Narratives</b>	Use or adapt existing narrative themes, where narratives are the baseline stories of a target audience. Narratives form the bedrock of our worldviews. New information is understood through a process firmly grounded in this bedrock. If new information is not consistent with the prevailing narratives of an audience, it will be ignored. Effective campaigns will frame their misinformation in the context of these narratives. Highly effective campaigns will make extensive use of audience-appropriate archetypes and meta-narratives throughout their content creation and amplification practices.
<b>Reframe Context</b>	Reframing context refers to removing an event from its surrounding context to distort its intended meaning. Rather than deny that an event occurred, reframing context frames an event in a manner that may lead the target audience to draw a different conclusion about its intentions.
<b>Flooding the Information Space</b>	Flooding and/or mobbing social media channels feeds and/or hashtags with excessive volume of content to control/shape online conversations and/or drown out opposing points of view. Bots and/or patriotic trolls are effective tools to achieve this effect.
<b>Trolls Amplify and Manipulate</b>	Use trolls to amplify narratives and/or manipulate narratives. Fake profiles/sock puppets operating to support individuals/narratives from the entire political spectrum (left/right binary). Operating with increased emphasis on promoting local content and promoting real Twitter users generating their own, often divisive political content, as it's easier to amplify existing content than create new/original content. Trolls operate wherever there's a socially divisive issue (issues that can/are to be politicized).
<b>Comment or Reply on Content</b>	Delivering content by replying or commenting via owned media (assets that the operator controls).
<b>Manipulate Platform Algorithm</b>	Manipulating a platform algorithm refers to conducting activity on a platform in a way that intentionally targets its underlying algorithm. After analyzing a platform's algorithm (see: Select Platforms), an influence operation may use a platform in a way that increases its content exposure, avoids content removal, or otherwise benefits the operation's strategy. For example, an influence operation may use bots to amplify its posts so that the platform's algorithm recognizes engagement with operation content and further promotes the content on user timelines.

## Executive Summary

The digital realm has witnessed TikTok's rapid ascension, captivating a global audience with its vibrant content and advanced algorithms. Yet, this rise has been shadowed by significant controversies, especially regarding debates in the United States over a potential TikTok ban, fueled by concerns over national security, data privacy, and the spread of misinformation. This contentious issue has sparked extensive discussions among lawmakers, technology experts, and digital communities, uncovering a complex web of digital manipulation and misinformation. From January 1 to March 25, 2024, an in-depth analysis recorded the involvement of 9,080 troll accounts in these debates, accounting for 11.73% of the total dialogue, thus underscoring the significant impact of troll-driven narratives on shaping public discourse.

This conversation spans three critical incidents: legislative efforts to restrict minors' access to social media, the Biden campaign's strategic engagement with TikTok, and debates surrounding the enactment of H.R. 7521. Each of these scenarios has ignited varied online reactions, with a notable share of the conversation being influenced by troll accounts. For instance, the initiative in Florida to limit social media access for minors saw 12.25% of its discussion driven by troll accounts, highlighting debates on the balance between individual rights and governmental authority. The discourse concerning the Biden campaign's use of TikTok and the legislative debates on H.R. 7521 further delve into issues of free speech, privacy, and governmental oversight, along with critiques of political leadership and representation. Trolls have extended their reach to manipulate discussions on a broad array of events, from global conflicts to international diplomacy these includes: International conflicts within European Union countries, notably Germany, Lithuania, and Sweden; the arrest of a Japanese crime boss involved in an attempt to smuggle nuclear materials to Iran; the Israel-Hamas conflict; the Ukrainian-Russian War; issues pertaining to China's international diplomacy and geopolitical strategies, each aiming to influence public opinion and policy.

The narrative varies across different platforms—Twitter, YouTube, Weibo, and TikTok—with observed very little manipulation activity on Facebook. Discussions on Twitter often revolve around political dissatisfaction and concerns over privacy and national security, while YouTube critiques focus on U.S. leadership and TikTok's content moderation practices. Weibo users tend to criticize U.S. policies, portraying them as bullying, whereas TikTok discussions emphasize speech restrictions and systemic critiques. These discussions often serve to challenge authority, question leadership, mobilize youth opposition, and notably, accuse the U.S. of violating First Amendment rights.

A comparison of narratives on American versus Chinese-owned social platforms reveals distinct focuses. Chinese platforms tend to argue that the U.S. approach to banning TikTok differs from that of other Western countries, suggesting that such a ban does not reflect the will of the American people and often pointing out that U.S. companies engage in more surveillance of their citizens than TikTok.

Furthermore, the analysis highlights a deliberate effort by troll accounts to echo narratives promoted by Chinese state-affiliated media, aiming to critique U.S. policies on free speech through the lens of the TikTok ban debate. By aligning with the viewpoints of outlets like Guangming Daily and Takungpao, these accounts play a pivotal role in spreading narratives that accuse the U.S. of hypocrisy regarding free speech and censorship, attempting to sway public opinion in favor of allowing TikTok to operate freely in the U.S. This concerted action underlines the strategic use of digital platforms in the broader geopolitical struggle, emphasizing the power of narrative in shaping the discourse on digital governance and international relations.

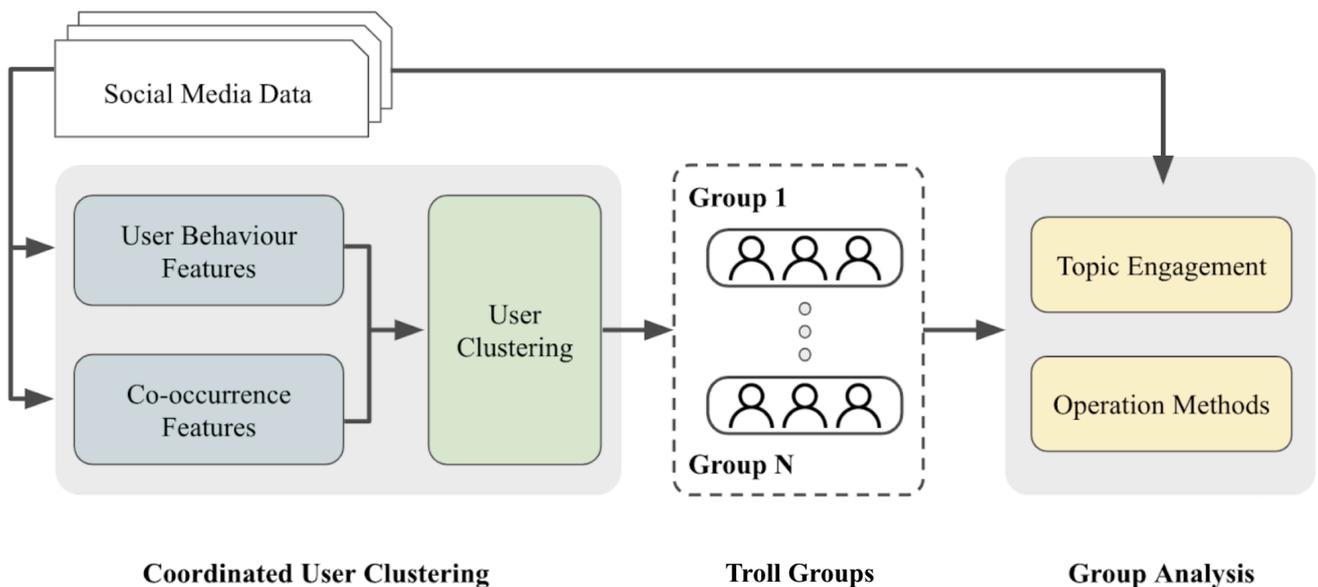
## Introduction

In the past few years, the digital arena has seen the explosive growth of TikTok, a platform that has enchanted millions with its compelling content and cutting-edge algorithms. Yet, this surge in popularity is intertwined with significant controversy. Central to the discord is the ongoing debate over the potential prohibition of TikTok in the United States, driven by apprehensions regarding national security, data privacy, and the proliferation of false information. This matter has ignited intense discussions among both policymakers and tech experts, drawing widespread attention across online communities. Amidst this chaos, thorough research has uncovered a complex environment where digital manipulation and misinformation thrive, revealing the intricate challenges at the heart of modern digital discourse.

## Methodology

Taiwan AI Labs utilizes our analytic tool "Infodemic" to conduct investigations into information operations across various social media platforms.

### Building Similarity Nodes Between User Accounts



Graph 1: An overview of the coordinated behavior analysis pipeline

Graph 1 illustrates the analysis pipeline of this report, consisting of three components:

- **User Feature Construction:** We evaluate and quantify the behavioral traits of users, converting these characteristics into user vectors for further analysis.
- **User Clustering:** Using these user vectors, we create a network of users with similar patterns and employ a community detection algorithm to pinpoint groups of users with high correlations, classifying them as collaborative units for additional scrutiny.

- **Group Analysis:** We explore the tactics and strategies of these collaborative units, focusing on their choice of subjects, methods of operation, and their inclination to either support or challenge certain entities.

## User Feature Construction

To capture user information on social forums effectively, we propose two feature sets:

### User Behaviour Features

Preparing data to highlight user behavior features is essential for deriving significant insights from the dataset, which includes a vast array of details pertaining to social media posts (or videos) and user interactions. We gathered a wide variety of raw social media data, subsequently converting it into a structured format with columns that depict various aspects of user behavior. This includes elements like the 'destination of user interactions' (indicated by `post_id` or `video_id`), the 'timing of user actions', and the 'domains of links shared by users', among others. These user behavior features will undergo further transformation and structuring to facilitate their use in assessing user similarity and for clustering purposes.

### Co-occurrence Features

Co-occurrence features aim to pinpoint users who often interact with similar topics or engage with identical articles. To quantify these features among users, we utilize Non-Negative Matrix Factorization (NMF), a mathematical method applied in data analysis and for reducing dimensionality. This technique decomposes a given matrix into two or more matrices, ensuring all elements within these matrices are non-negative.

## User Clustering

### User Similarity Evaluation

After establishing user features, we move to examine the coordinated relationships among users. For behavioral features, we conduct comparisons of various behaviors between pairs of users and scale the results to a range from 0 to 1. For example, regarding the timing of user activities, we document the hours of activity for each user over a week in a 7x24-dimensional matrix. Subsequently, we calculate the cosine similarity between user pairs based on their activity timing matrices.

In terms of co-occurrence features, cosine similarity is also employed to gauge the resemblance between users' co-occurring vectors. This involves calculating the cosine of the angle between these vectors to determine the degree of similarity in users' responses or actions. This method proves particularly effective in social media studies, enabling the grouping of users by shared behavioral patterns. Users exhibiting high cosine similarity are indicative of a closely coordinated behavior pattern, revealing clusters of users with similar interests or engagement habits.

## User Clustering

Once we've calculated pairwise similarities among users from their individual features, we proceed to connect user pairs that exhibit a similarity beyond a set threshold by establishing an edge between them, thus forming a user network. Following the creation of this network, we employ the Infomap algorithm to cluster it. Infomap is a community detection algorithm that identifies structures within networks based on the flow of information. Communities discovered within this network are subsequently classified as troll groups for further analysis in subsequent sections. This method allows us to systematically identify and categorize groups of users exhibiting coordinated behavior patterns, which are indicative of troll activity.

## Group Analysis

### Opinion Clustering

To effectively decipher the narratives put forth by each user group, we utilized a text clustering approach on the comments made by troll groups. By leveraging a pre-trained text encoder, we transformed each comment into vector form. We then employed a hierarchical clustering algorithm to organize similar posts into cohesive groups. These clustered groups of posts will be analyzed further in subsequent discussions, providing a structured framework to examine and understand the narratives and themes prevalent within troll group communications.

### Stance Detection and Narrative Summary

Large Pretrained Language Models have showcased their effectiveness in identifying entities within textual content and providing insightful explanations about them. This functionality aids in grasping the key components of discourse, especially in analyzing the influence of comments and evaluations on these recognized entities.

In our analysis, we utilize Taiwan LLM for text examinations. Taiwan LLM is a substantial language model that has been pre-trained on a corpus predominantly in the native Taiwanese language. It has demonstrated exceptional ability in understanding Traditional Chinese and is particularly adept at identifying and interpreting topics and entities related to Taiwan. Specifically, we employ Taiwan LLM to discern essential topics, entities, and names of organizations mentioned in each comment. Additionally, it evaluates the comment author's perspective towards these entities, classifying their sentiment as positive, neutral, or negative. This method is systematically applied across all clusters of opinions.

Ultimately, we aim to calculate the proportion of each primary topic or entity mentioned within the opinion groups, alongside the percentage of positive or negative sentiment linked with each. Moreover, we generate summaries for each opinion cluster using the language model, which assists data analysts in quickly comprehending the broad overview of the event and the prevailing sentiments within the discourse.

## Data Coverage

AI Labs utilized its proprietary Infodemic platform to analyze the recent TikTok banning event. The data analyzed for this study spans from January 1, 2024, to March 25, 2024. The research recorded 65 events and tracked 460 instances of media engagement. It was identified that 9,080 troll accounts

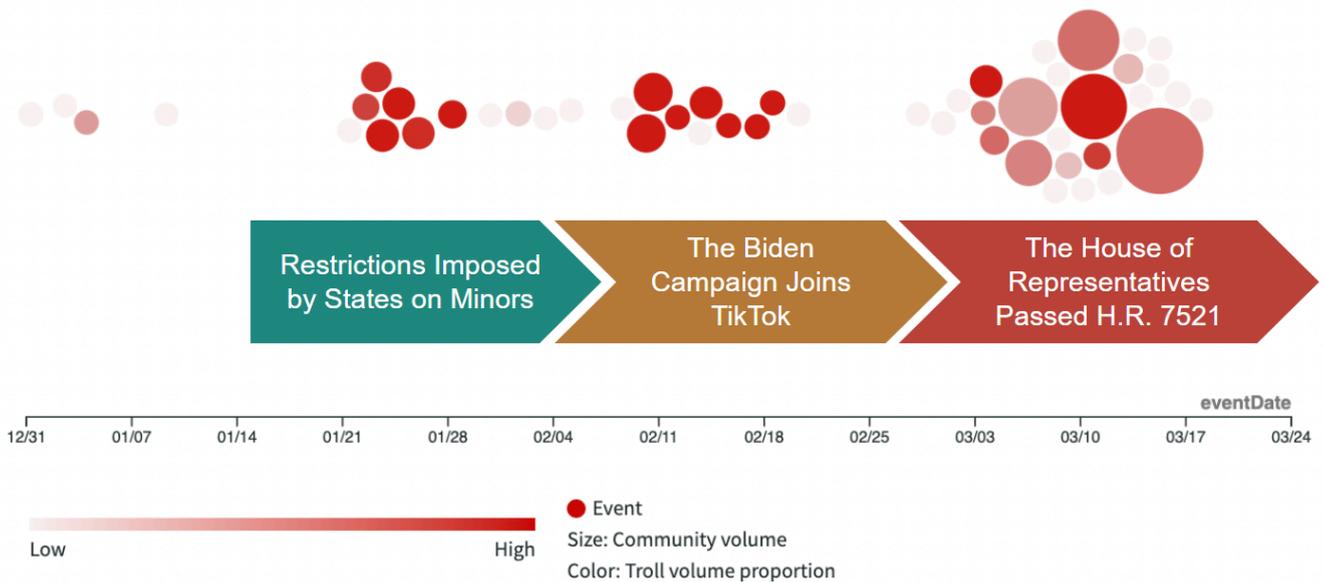
participated in the discourse across social media platforms. Out of the total 111,809 discussion volume, 11.73% was found to originate from troll accounts, highlighting their substantial influence in shaping online conversations.

Events	Media Volume	(PRC) State-affiliated Media (%)	Troll Accounts	Community Volume	Troll Volume (%)
65	460	5(1.09%)	9,080	111,809	13,119 (11.73%)

Table 1: Analyzed data quantity during the period of the the TikTok banning event (from <https://infodemic.cc>)

## Major Timeline and Most Representable Battlefields

The events analyzed by the Infodemic platform are presented in a BeeSwarm Plot, as shown in Graph 2. The observed events fall into three primary categories: restrictions imposed by states on minors, the Biden campaign joins tiktok, and the house of representatives passed H.R. 7521. Please note that due to the limited accessibility of TikTok data, the graphs may not accurately represent the full extent of troll manipulation volume.



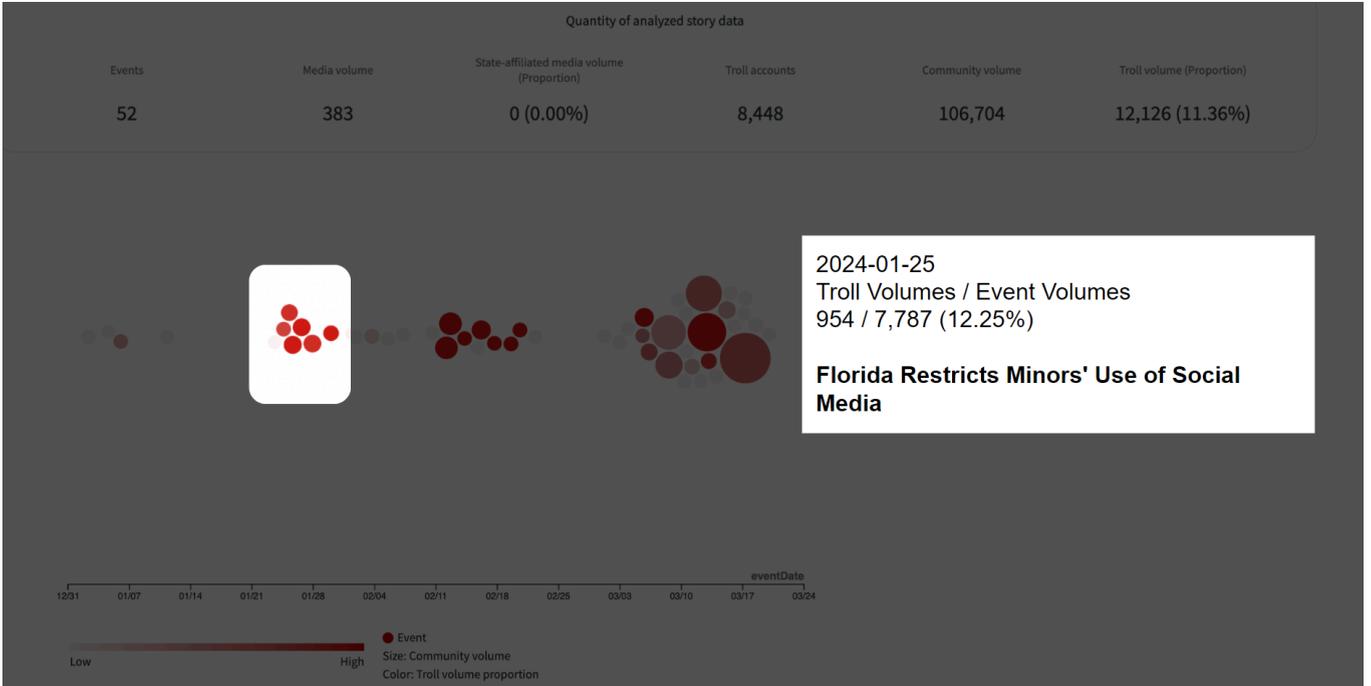
- \* Each circle represents an event related to this manipulated story
- \*\* The size of each circle defined by the sum of the social discussion of that event
- \*\*\* The darker the circle is, the higher the proportion of troll comments in the event

Graph 2: Event overview by timeline (from <https://infodemic.cc>)

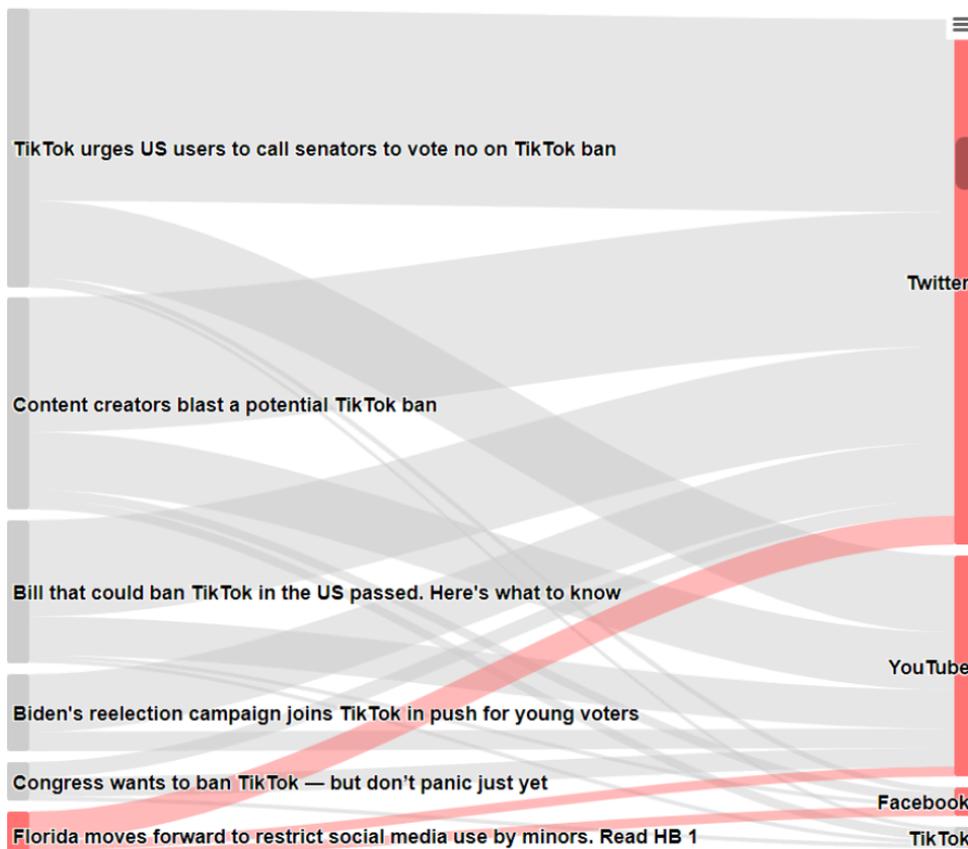
## 2024, January: Restrictions Imposed by States on Minors

Restrictions imposed by states on minors, predominantly centers around Florida's legislative move to limit minors' access to social media. This development, commencing on January 25, 2024, sparked

considerable discussion, with 12.25% of the discourse attributable to troll accounts. Analysis, as illustrated in the Sankey diagram, reveals that a significant portion of this troll-generated conversation originated on Twitter. The narratives predominantly revolve around the tension between individual rights and the state government's authority, the implications of restricting information and how labor laws pertain to minors, and broader considerations regarding children's rights.



Graph 3: Event overview by timeline (from <https://infodemic.cc>)



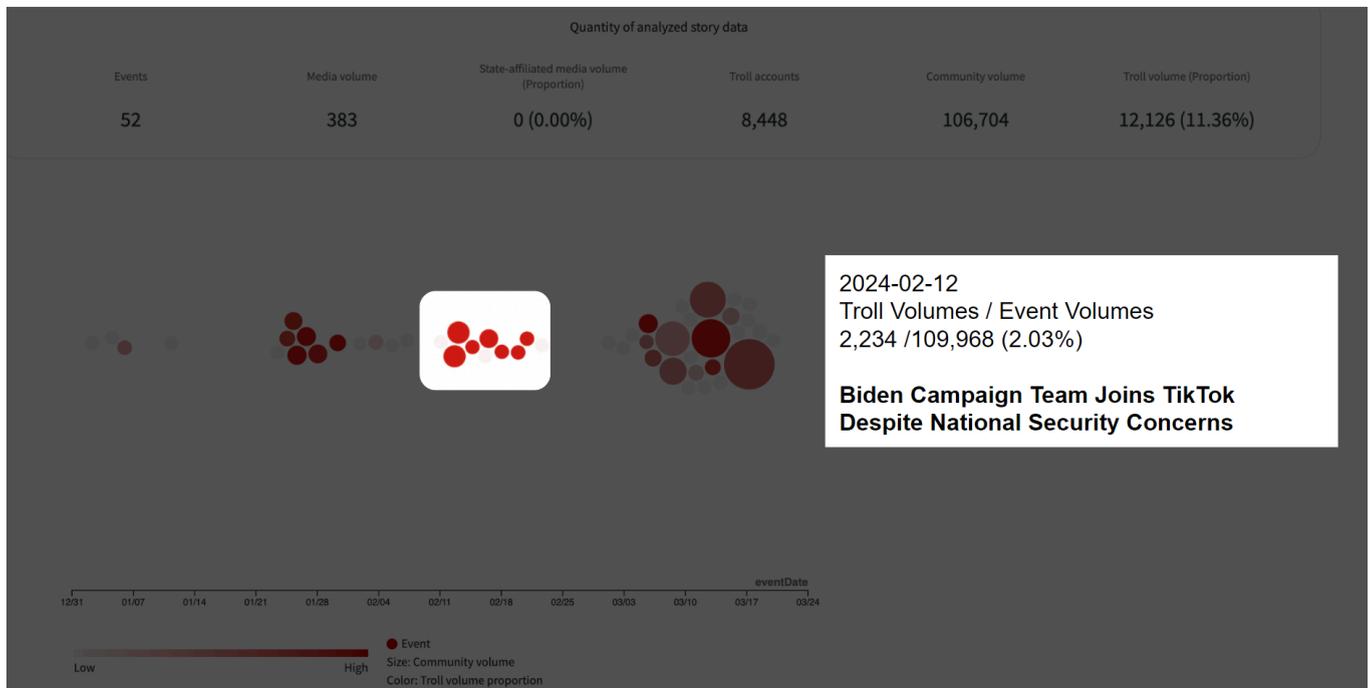
Graph 4: Sankey diagram illustrating the distribution of troll activity across various events to social media platforms. (from <https://infodemic.cc>)

Percentage	Narratives
6.3%	The attempt by Florida to ban TikTok has sparked a broader debate about the extent to which state governments can implement drastic measures in the name of public safety. This situation underscores the ongoing conflict between individual rights and the authority of state governments.
3%	Concerns about kids working instead of using social media were raised. Criticism is directed towards motives related to restricting information and labor laws affecting minors.
2.8%	Concerns are raised over possible consequences for children's rights.

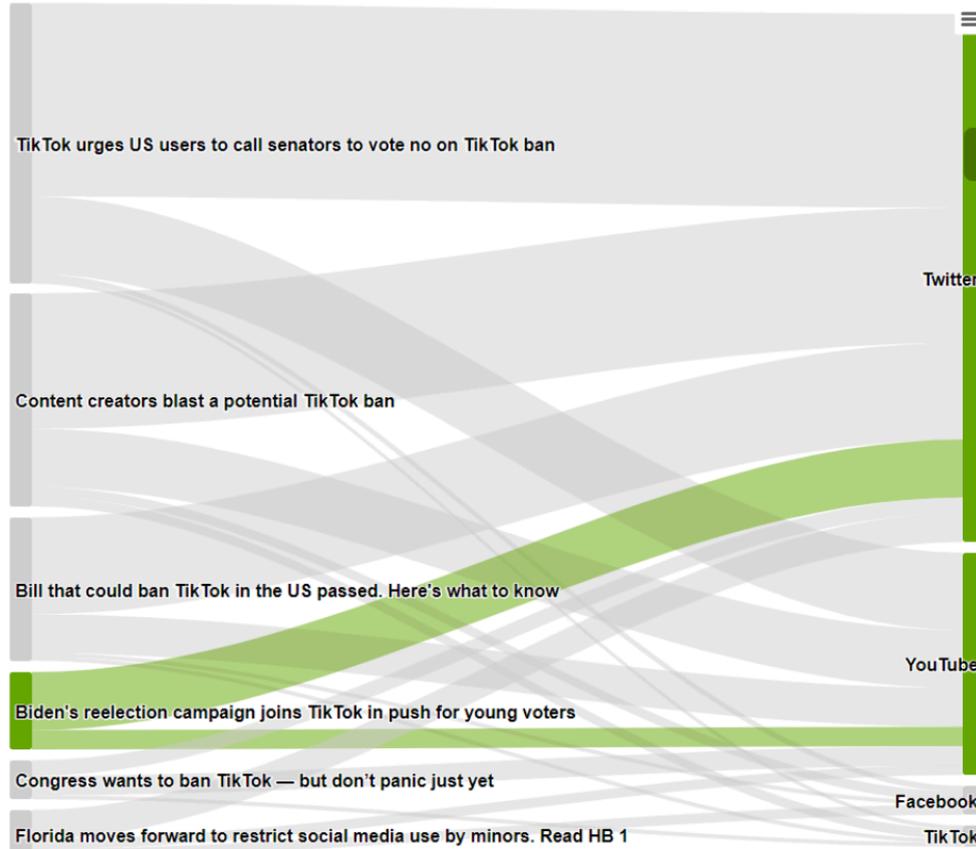
Table 2: Analyzed data quantity during the period of the the TikTok banning event (from <https://infodemic.cc>)

### 2024, February: The Biden Campaign Joins TikTok

Regarding the Biden campaign's engagement with TikTok, the spotlight is primarily focused on the event titled "Biden Campaign Team Joins TikTok Despite National Security Concerns." This strategic move, initiated on February 12, 2024, saw a relatively lower troll discussion volume, accounting for 2.03% of the overall discourse. According to insights drawn from the Sankey diagram, the majority of troll-originated conversations took place on Twitter, with a significant presence on YouTube as well. The discussions driven by these trolls were characterized by narratives questioning Joe Biden's suitability for presidency and unfounded allegations of bribery linked to China.



Graph 5: Events overview by timeline (from <https://infodemic.cc>)



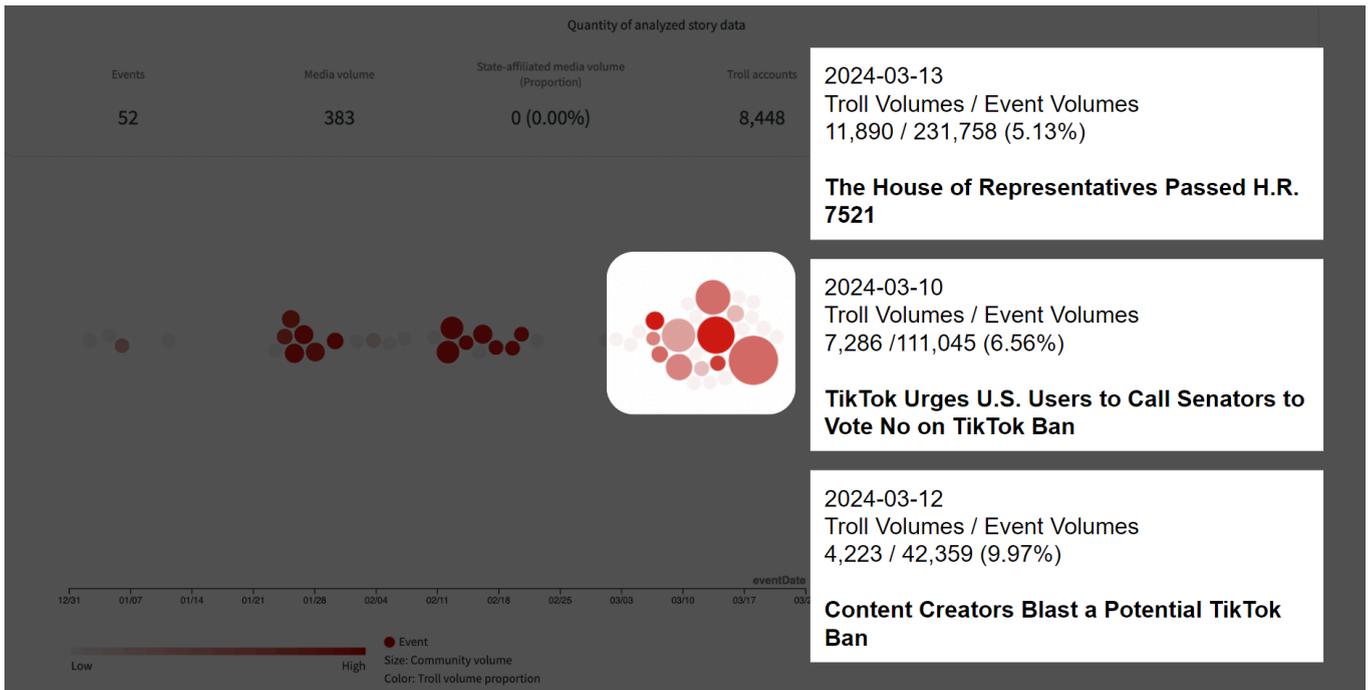
Graph 6: Sankey diagram illustrating the distribution of troll activity across various events to social media platforms. (from <https://infodemic.cc>)

Percentage	Narratives
9.5%	Critics on social media accuse Biden of being unfit and suggest he should be in a nursing home, Gitmo, or a Gulag.
5.2%	Comments have been made implying baseless associations between the Democratic Party, Biden, and China involving bribery.

Table 3: Analyzed data quantity during the period of the the TikTok banning event (from <https://infodemic.cc>)

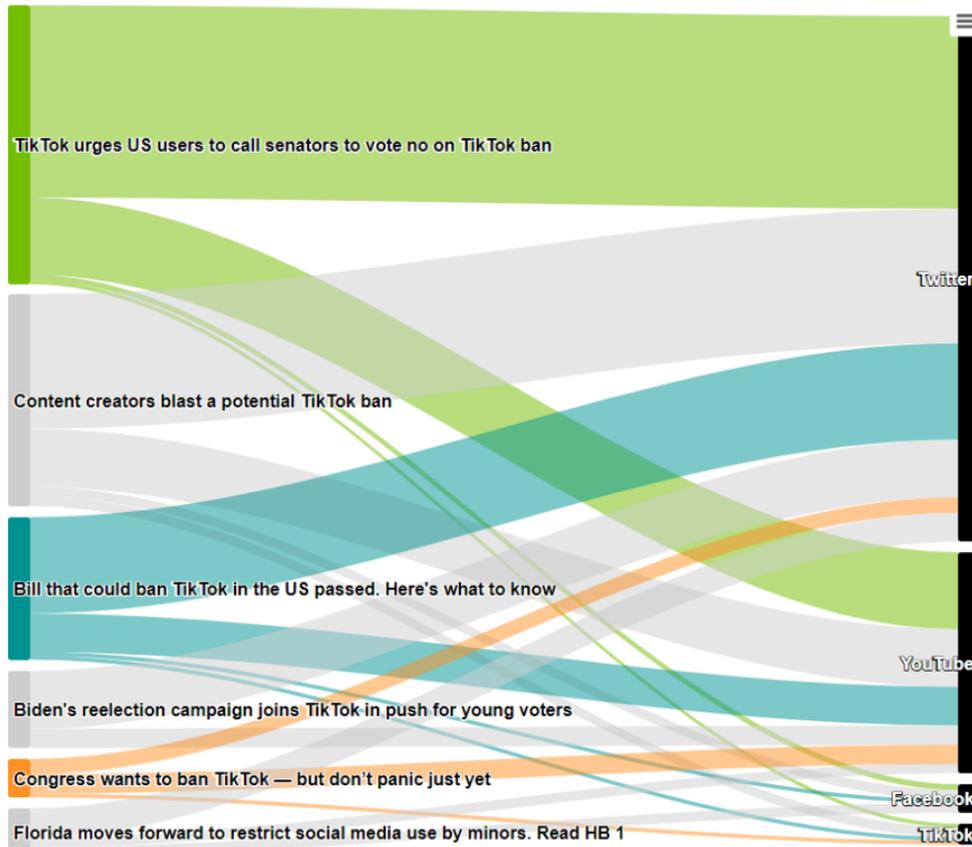
## 2024, March: The House of Representatives Passed H.R. 7521

The third category delves into the legislative process, particularly the passage of H.R. 7521 by the House of Representatives, and encompasses three significant events. These events are: the passing of H.R. 7521, observed with a troll discussion volume of 5.13%; TikTok's appeal to U.S. users to contact Senators to oppose the TikTok ban, which saw a 6.56% contribution from troll accounts; and content creators voicing their opposition to a potential TikTok ban, generating the highest troll volume at 9.97%. These discussions gained momentum around March 10, 2024, and spanned across multiple platforms, including Twitter, YouTube, Weibo, and TikTok, with the bulk of troll-driven conversations occurring on Twitter and YouTube.



Graph 7: Events overview by timeline (from <https://infodemic.cc>)

Within the event where TikTok encouraged U.S. users to lobby senators against the TikTok ban and the event the house of representatives passed H.R. 7521, the discourse unfolded along three distinct troll-generated narratives. First, there was a strong argument positing that the bill represented an infringement on freedom of speech, suggesting that the proposed legislation was a form of censorship. Secondly, discussions highlighted concerns about government involvement in private enterprises, reflecting apprehensions regarding the extent to which the state should interfere in the digital and business realms, as well as debating the benefits individuals gain from their interactions with such platforms. Lastly, a narrative of frustration over political representation emerged, with users expressing dissatisfaction with how politicians understand and represent their interests, especially in relation to digital freedoms and the online ecosystem.

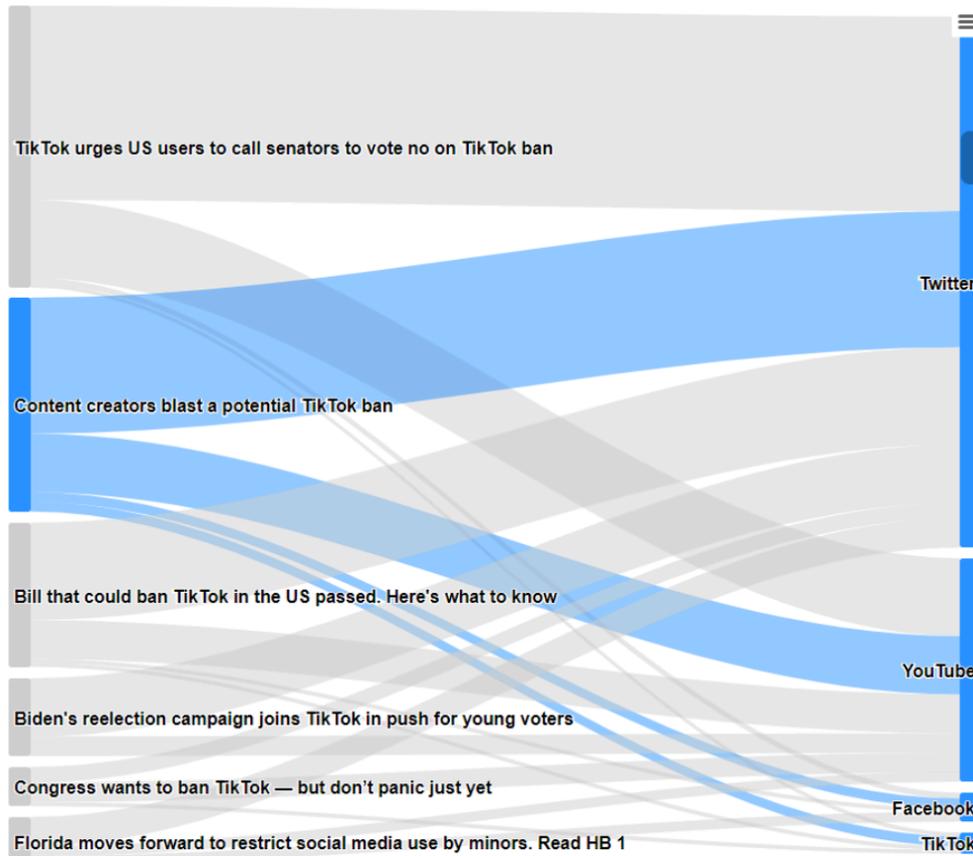


Graph 8: Sankey diagram illustrating the distribution of troll activity across various events to social media platforms. (from <https://infodemic.cc>)

Percentage	Narratives
8.1% 	The bill under discussion is controversial due to concerns about penalizing users and censoring speech.
8%	The commentary encompasses discussions on the role of government involvement in private enterprises and the individual advantages derived from such interactions
6.4%	Critics on social media express frustration over politicians' representation.

Table 4: Analyzed data quantity during the period of the the TikTok banning event (from <https://infodemic.cc>)

Under the event where content creators expressed strong opposition to a potential TikTok ban, the discourse primarily revolved around two interrelated troll-generated narratives. Firstly, there were significant concerns regarding privacy, government surveillance, and the specter of potential censorship. This narrative suggests apprehension about the overreach of governmental authority into personal freedoms and the digital domain, pointing to fears that a ban could set a precedent for future regulatory actions that might infringe upon privacy rights and the ability to freely exchange ideas. Secondly, the discourse strongly emphasized threats to freedom of speech. This perspective underscores the belief that a TikTok ban would not only limit content creation and sharing but also diminish a crucial platform for expression and communication.



Graph 9: Sankey diagram illustrating the distribution of troll activity across various events to social media platforms. (from <https://infodemic.cc>)

Percentage	Narratives
13.7%	Concerns about privacy, government surveillance, potential censorship.
4.5% 	Threats to free speech rights.

Table 5: Analyzed data quantity during the period of the the TikTok banning event (from <https://infodemic.cc>)

## Narratives on the legislative process surrounding H.R. 7521's passage by troll users by platforms

The discourse surrounding the passage of H.R. 7521 by the House of Representatives, as reflected across Twitter, YouTube, Weibo, and TikTok, illustrates a wide spectrum of narratives from the troll users.

### Twitter Narratives

On Twitter, the conversation is marked by a broad array of concerns including dissatisfaction with the political establishment, perceived efforts by the U.S. to stoke anti-China sentiments under the guise of TikTok discussions, and debates over data privacy, censorship, and national security implications.

There's also a notable narrative suggesting that Israel's influence might have played a role in the decision to target TikTok, introducing an international diplomatic dimension to the discourse.

Percentage	Narratives
8.1%	Comments touch on frustration with political parties, criticism of the US influencing anti-China sentiments through TikTok, and pessimism about the country's future.
8%	Social media users are discussing the implications of the new bill, raising concerns about data sharing, censorship powers, and national security issues regarding TikTok.
6.4%	Comments suggest that banning TikTok could benefit Israel, implying that Israel may have influenced the decision to ban TikTok.

Table 6: Analyzed data quantity during the period of the the TikTok banning event (from <https://infodemic.cc>)

## Youtube Narratives

YouTube discussions veer into criticisms of American leadership and the country's labor laws, alongside skepticism regarding the effectiveness and fairness of TikTok's content moderation practices. These narratives suggest a critical examination of broader societal and governance issues beyond the app itself, reflecting YouTube's role as a platform for in-depth analysis and commentary.

Percentage	Narratives
8.1%	The comments criticize America's leadership, question support for morally questionable figures, highlight weak labor laws.
6.4%	Skepticism about the TikTok's content screening.

Table 7: Analyzed data quantity during the period of the the TikTok banning event (from <https://infodemic.cc>)

## Weibo Narratives

On Weibo, the narratives pivot towards accusing the U.S. of international bullying, specifically targeting Chinese companies like ByteDance, and discussions on the perceived U.S. attempts to exert control over foreign corporations and assets through legislative means. Conversations also touch on the role of Jewish organizations in the legislative lobbying process, highlighting perceived external influences on U.S. policies.



Graph 10: Sankey diagram illustrating the distribution of troll activity across various events to social media platforms. (from <https://infodemic.cc>)

Percentage	Narratives
9.4%	Accusing the U.S. of bullying other nations and companies, especially ByteDance (TikTok), and trying to legalize control over foreign firms and assets through legislation.
8%	Discussing sanctions the U.S. government impose on TikTok, including bans and forced divestment.
4.5%	This initiative is led by the Jewish Federation of North America, with over a hundred Jewish organizations lobbying outside the legislature.

Table 8: Analyzed data quantity during the period of the the TikTok banning event (from <https://infodemic.cc>)

### TikTok Narratives

TikTok users focus on the implications of the bill for speech restrictions, privacy, and internet freedom, alongside critical views on the motivations behind American political actions and policies. There's also a distinct narrative criticizing the Republican party and expressing dismay at the overall situation, indicating a personalized and emotional response from the platform's user base.



Graph 11: Sankey diagram illustrating the distribution of troll activity across various events to social media platforms (from <https://infodemic.cc>)

Percentage	Narratives
19.3% 	Comments express concerns over speech restrictions, privacy, and fears of a TikTok ban, urging for internet freedom.
11.3%	These comments reflect skepticism towards U.S. domestic policies and the intentions behind the American political system and policies.
4.8%	The comments criticize Republicans and express heartbreak over the situation.

Table 9: Analyzed data quantity during the period of the the TikTok banning event (from <https://infodemic.cc>)

### Trolls Echo PRC State-affiliated Media

The specific narrative on freedom of speech, resonating across various social platforms, finds support in the perspectives shared by China-affiliated media outlets Guangming Daily and Takungpa.

Guangming Daily emphasizes TikTok's stance that the House of Representatives' decision marks merely the beginning of an extensive process rather than its culmination. The platform expresses its intention to continue efforts to influence the Senate, urging its user base to reach out to senators in defense of the freedom of speech rights enshrined in the First Amendment of the U.S. Constitution.

Takungpao, referencing a report by "The Wall Street Journal," critiques the recent stringent actions of the U.S. government against foreign entities like TikTok as reflective of a shift towards greater nationalism and protectionism. Such a stance, the report suggests, could deter foreign investment and raise apprehensions among international businesses. Furthermore, it highlights the potential legal controversies that might ensue from measures to force TikTok's sale or enact a ban, particularly concerning possible infringements on free speech rights as guaranteed by the First Amendment.

## Main Troll Groups

During the TikTok banning event, Taiwan AI Labs observed unusual activity from three troll groups. These groups generated significant noise and manipulation on both X and YouTube platforms, actively engaging in international political controversies and policy-related issues without confining themselves to any single nation. We believe the behavior of these three troll groups is atypical and warrants further investigation. Below is a summary of information related to these troll groups.

### Troll Group: Twitter#2774

Twitter #2774, observed by AI Labs, is the most active troll group on X (formerly Twitter), with 177 accounts participating in 3,163 stories.

Troll Accounts	Operated stories	Target entities
177	3,163	1,907

Table 10: Summary of Twitter #2774 (from <https://infodemic.cc/collab/twitter@2774>)

### Abnormal Behaviors

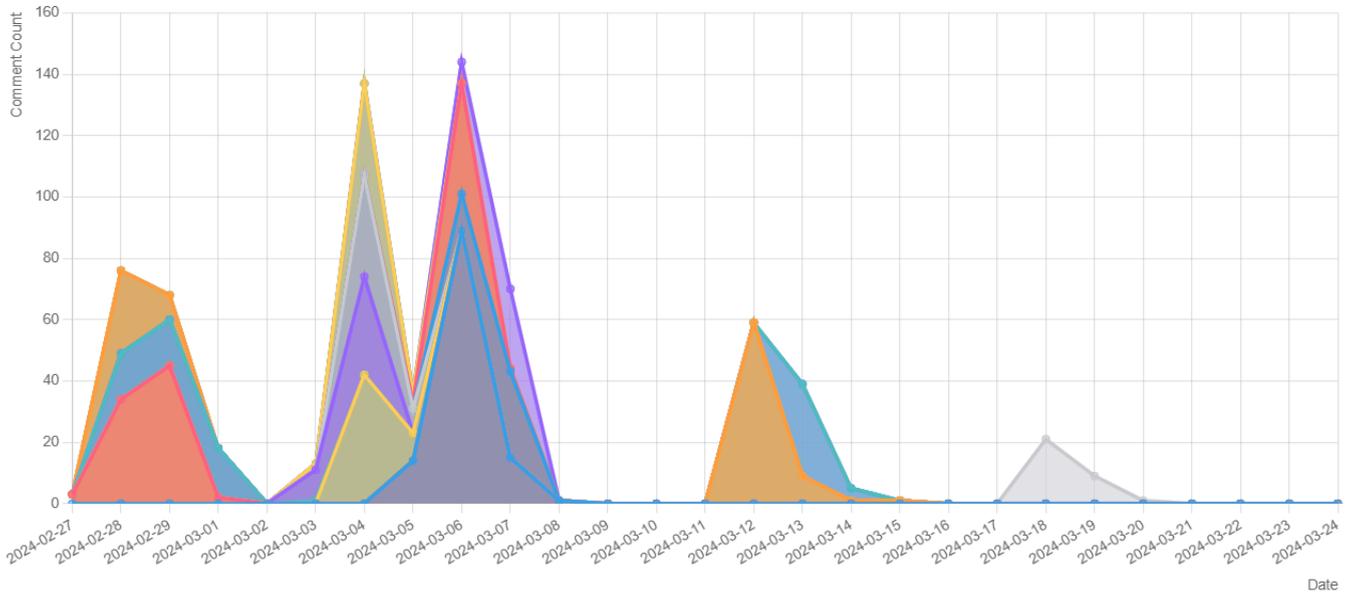
Observing the activity trend of X troll group #2774 over the past three months, it's noteworthy that the volume of comments is significantly higher compared to regular users



### Operated Stories

The Twitter troll group #2774 actively engages in global conflict-related issues, including the Israel-Hamas conflict, the Ukrainian-Russian War, as well as international conflicts involving European

Union countries such as Germany, Lithuania, and Sweden. They even delve into cases such as the arrest of a Japanese crime boss for attempting to smuggle nuclear materials to Iran.



Graph 12: Operated stories of Twitter #2774 by timeline (from <https://infodemic.cc/collab/twitter@2774>)

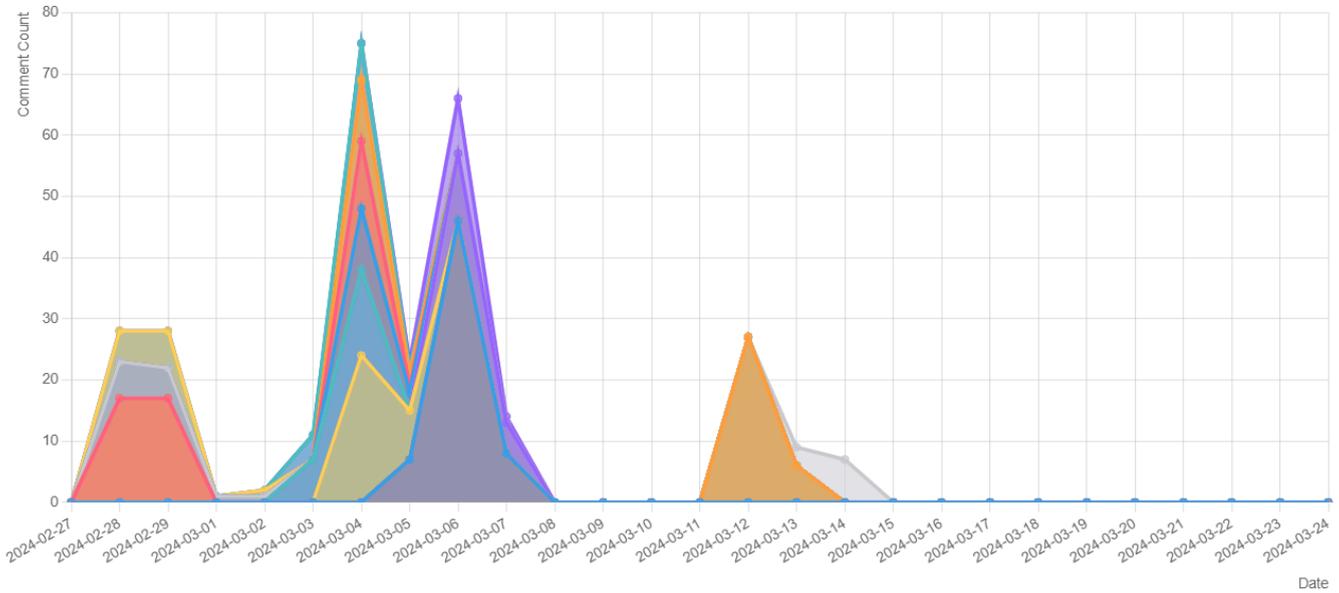
Event time (UTC+8)	Title	Community volume	Troll volume (%)
2024/3/5 19:10 2024/3/9 17:01	Nikki Haley drops out of 2024 race, doesn't endorse Trump for GOP presidential nomination	142,487	119 (4.55%)
2024/2/24 00:50 2024/3/1 05:30	Supreme Court to review Trump immunity	114,548	84 (3.21%)
2024/3/11 18:00 2024/3/13 03:44	4 key questions ahead of Robert Hurs testimony on Bidens mishandling of classified documents	116,029	70 (2.68%)
2024/3/4 16:18 2024/3/5 03:45	Supreme Court keeps Trump on Colorado ballot, rejecting 14th Amendment push	98,166	51 (1.95%)
2024/3/1 00:54 2024/3/2 06:03	Trump again presses for delay of classified documents trial until 2025	39,686	47 (1.80%)

Table 11: Top 5 operated stories of Twitter #2774 (from <https://infodemic.cc/collab/twitter@2774>)



### Operated Stories

The Twitter troll group #2774 actively engages in issues related to the Israel-Hamas conflict, the Ukrainian-Russian War, as well as topics concerning China's international diplomacy and geopolitical affairs.



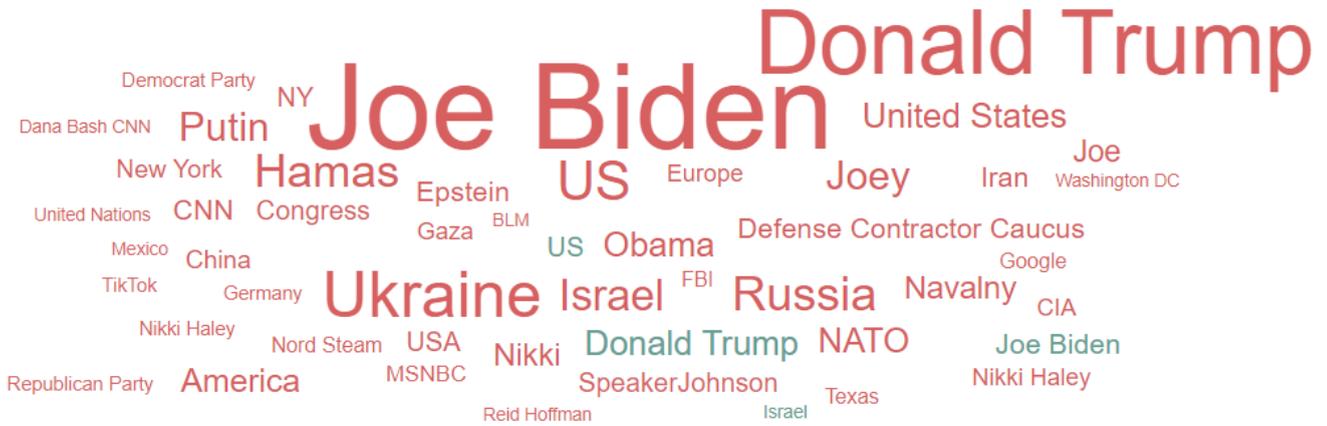
Graph 14: Operated stories of Twitter #6257 by timeline (from <https://infodemic.cc/collab/twitter@6257>)

Event time (UTC+8)	Title	Community volume	Troll volume (%)
2024/3/5 19:10 2024/3/9 17:01	Nikki Haley drops out of 2024 race, doesn't endorse Trump for GOP presidential nomination	142,487	61 (6.01%)
2024/2/24 00:50 2024/3/1 05:30	Supreme Court to review Trump immunity	114,548	34 (3.35%)
2024/3/11 18:00 2024/3/13 03:44	4 key questions ahead of Robert Hurs testimony on Bidens mishandling of classified documents	116,029	33 (3.25%)
2024/3/4 16:18 2024/3/5 03:45	Supreme Court keeps Trump on Colorado ballot, rejecting 14th Amendment push	98,166	32 (3.15%)
2024/3/3 18:00 2024/3/4 11:03	Nikki Haley defeats Trump in Washington DC for first primary win	60,707	21 (2.07%)

Table 13: Top 5 operated stories of Twitter #6257 (from <https://infodemic.cc/collab/twitter@6257>)

### Targets of Troll Activities

■ Negative   
 ■ Positive   
 大, 小 Font size represents troll volume



Graph 15: Troll activity targets of Twitter #6257 (from <https://infodemic.cc/collab/twitter@6257>)

### Troll Group: YouTube #253

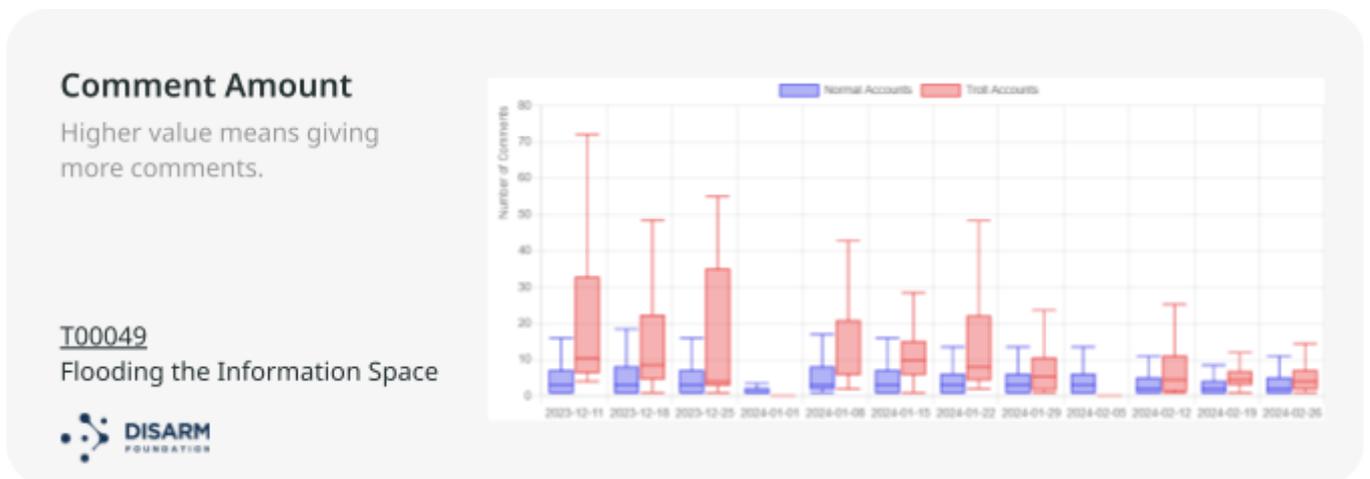
YouTube #253 is one of the most active troll groups on the YouTube platform, comprising 1,503 accounts and participating in 1,748 stories.

Troll Accounts	Operated stories	Target entities
1,503	1,748	2,783

Table 14: Summary of YouTube #253 (from <https://infodemic.cc/collab/youtube@253>)

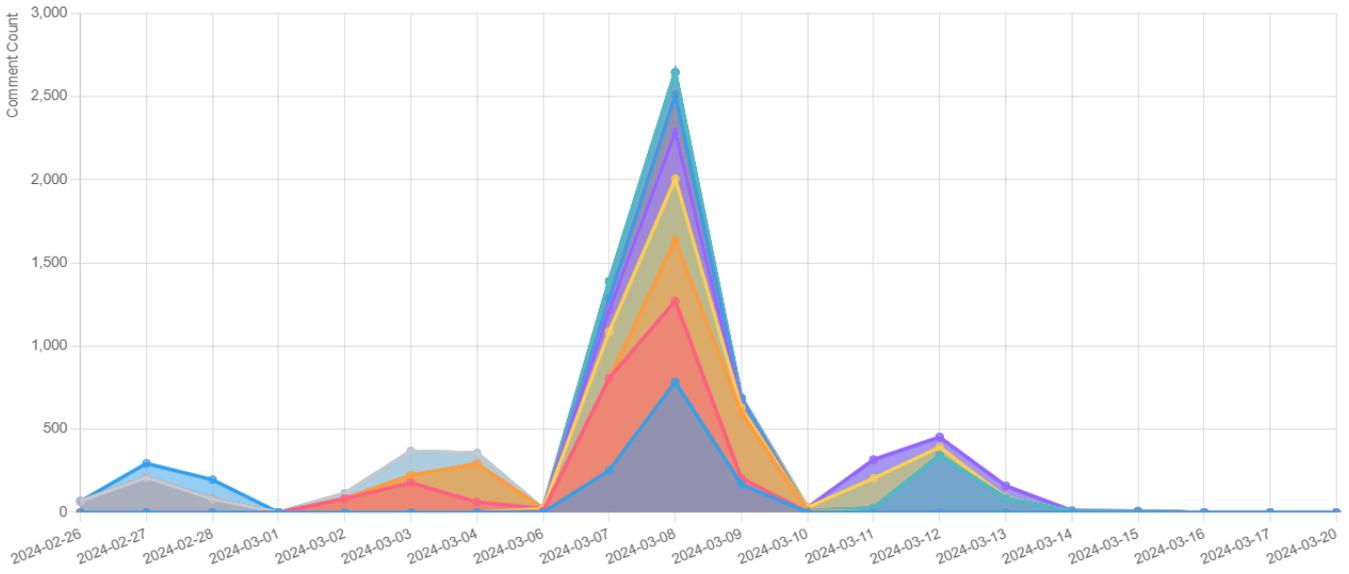
### Abnormal Behaviors

Observing the activity trend of YouTube troll group #253 over the past three months, it's noteworthy that the volume of comments is significantly higher compared to regular users.



### Operated Stories

The YouTube troll group #253 actively engages in issues related to participating in the US presidential election, discussing various conspiracy theories, and making comparisons between the treatment of January 6th demonstrators and President Biden's handling of classified documents. They express frustration with perceived injustices and claim a cover-up in the legal system. Some show support for Biden's actions, citing his role as the chief executive. They also participate in domestic events in China, such as the northern China gas explosion case.

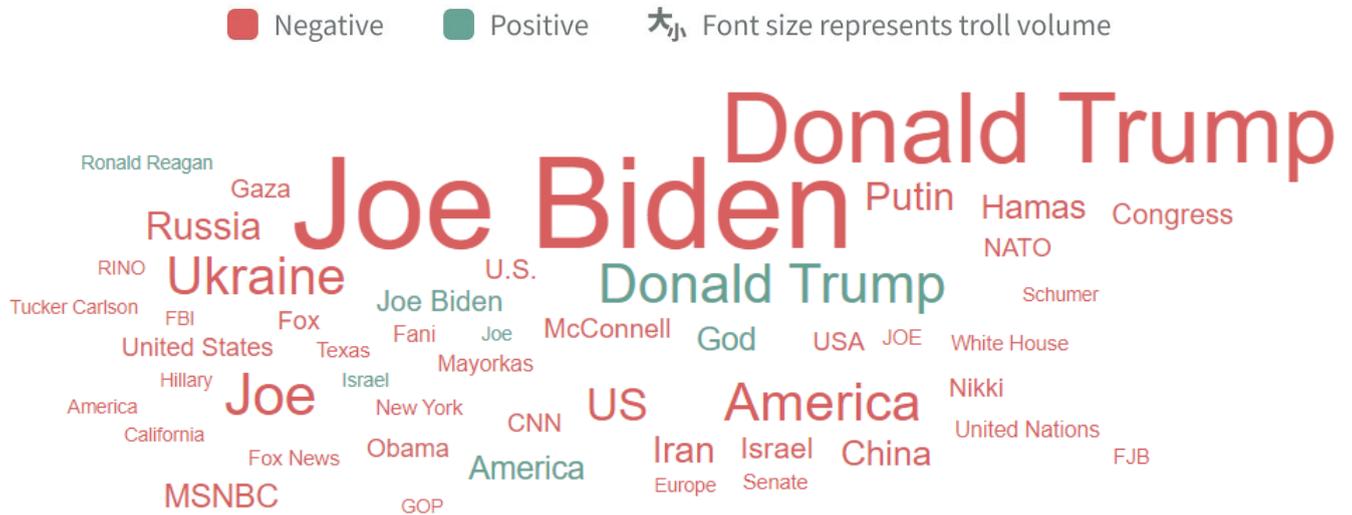


Graph 16: Operated stories of YouTube #253 by timeline(from <https://infodemic.cc/collab/youtube@253>)

Event time (UTC+8)	Title	Community volume	Troll volume (%)
2024/3/8 12:44 2024/3/9 11:53	Biden Kate Cox, Texas abortion ban in State of the Union. Read what he said.	223,954	1,207 (6.64%)
2024/3/6 23:00 2024/3/8 01:00	What to know about Kate Cox: Biden State of the Union guest to spotlight abortion bans	196,570	1,088 (5.99%)
2024/3/9 12:35 2024/3/10 16:55	Biden Expresses Regret for Calling an Undocumented Immigrant 'an Illegal'	132,210	770 (4.24%)
2024/3/8 18:07 2024/3/8 18:07	Sean Hannity Deploys New Nickname For Joe Biden And Democrats Actually Love It	157,523	688 (3.79%)
2024/3/11 18:00 2024/3/13 03:44	4 key questions ahead of Robert Hurs testimony on Bidens mishandling of classified documents	116,029	468 (2.57%)

Table 15: Top 5 operated stories of YouTube #253 (from <https://infodemic.cc/collab/youtube@253>)

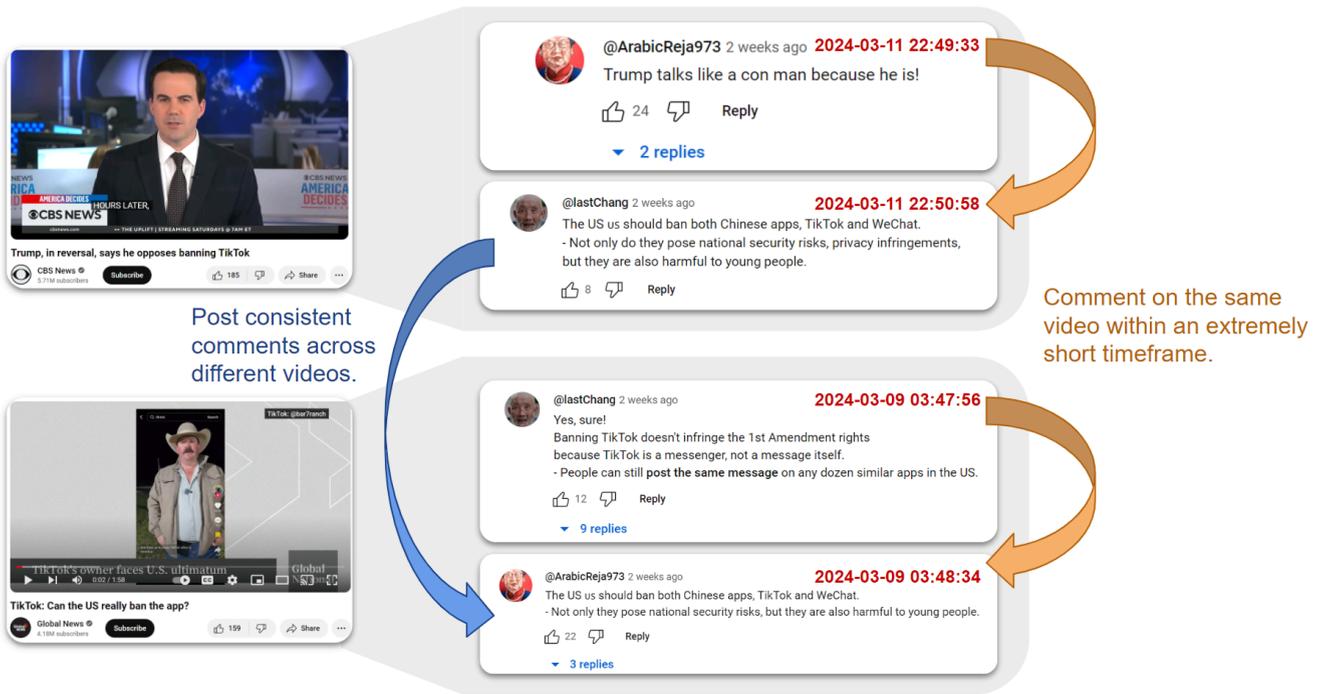
### Targets of Troll Activities



Graph 17: Troll activity targets of YouTube #253 (from <https://infodemic.cc/collab/youtube@253>)

### Operational Examples of Troll Groups

Exemplifying the operations of troll groups on YouTube, members collaborate to rapidly comment on the same video within an extremely short time frame, thereby artificially inflating the engagement metrics of the video and subsequently boosting its click-through rate. Additionally, members of these groups also post identical comments across different videos, aiming to proliferate the intended narrative orchestrated by the organization.



Graph 18: Cognitive Manipulation Techniques Example for YouTube

## DISARM Techniques Used by Troll Groups

Regarding the DISARM framework<sup>2</sup> from NATO, the researcher found that troll group operations on Facebook, YouTube, PTT, and TikTok were divided into two phases: Prepare and Execute

Phase	Tactic	Twitter	Youtube	Weibo	Tiktok
Prepare	T0003 Leverage Existing Narratives	○	○	◎	○
Execute	T0023.001 Reframe Context	◎	◎	◎	◎
	T0049 Flooding the Information Space	◎	◎	○	◎
	T0049.001 Trolls Amplify and Manipulate	◎	○	◎	○
	T0116 Comment or Reply on Content	◎	◎	○	◎
	T0121 Manipulate Platform Algorithm	◎	○	◎	◎

- indicates observed manipulative behaviors that align with this Tactic.
- ◎ signifies observed manipulative behaviors that very closely match this Tactic.

Table 16: DISARM Tactics used on each platform

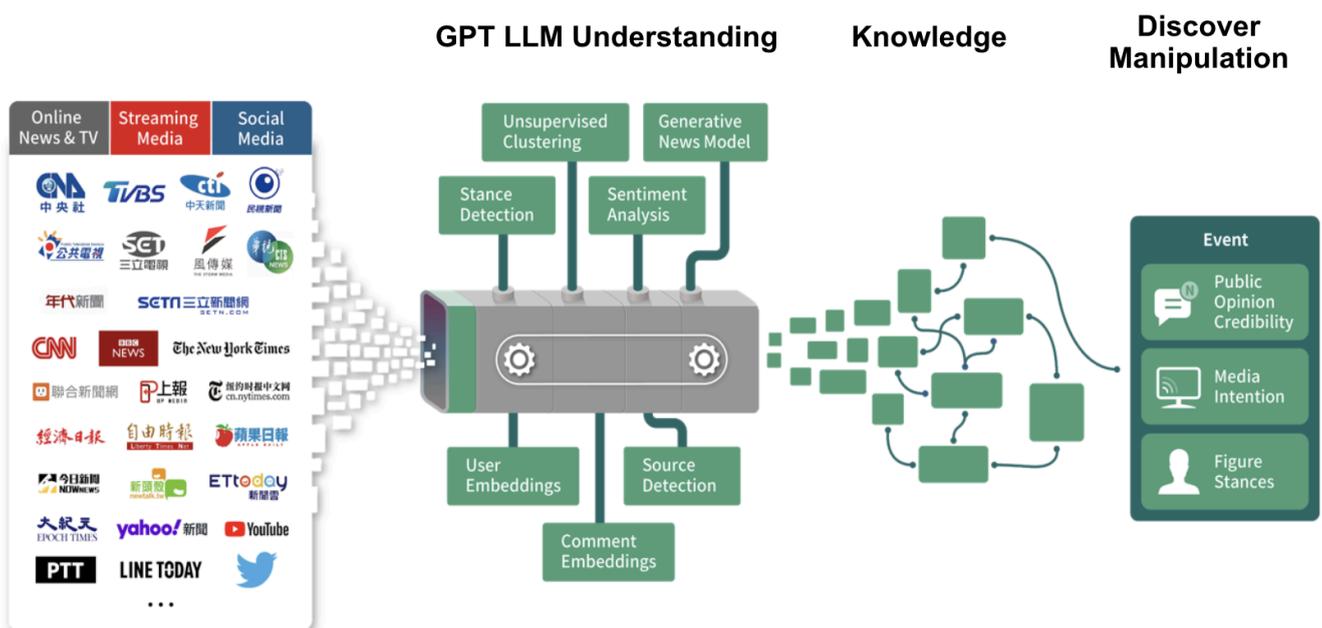
<sup>2</sup> DISARM Disinformation Analysis and Risk Management is an open-source framework designed to describe and understand the behavior parts of FIMI/disinformation. It sets out best practices for fighting disinformation through sharing data & analysis and can inform effective action. The Framework has been developed, drawing on global cybersecurity best practices. <https://www.disarm.foundation/>

## The Infodemic Platform

During the pandemic, Taiwan AI Labs collaborated internationally to develop trustworthy and responsible AI in healthcare while addressing the global challenge of misinformation related to the pandemic. Working with global partners, we established mechanisms to detect such activities. Taiwan AI Labs initially used AI to observe and understand the behavior of various accounts, identifying coordinated activities to detect synchronized accounts.

Troll accounts are defined as a group of accounts not operated by genuine users. These could be accounts publishing specific content as per official directives, or those controlled programmatically or through PR firms, disseminating particular narratives in a non-organic, organized manner. By leveraging generative technologies and large language models (LLMs), Taiwan AI Labs analyzed billions of social media activities to unearth over 30,000 troll groups, understanding the content and patterns of their operations across more than two million topics. This helps to uncover the targets, methods, and possible motives behind these operations.

With the growing global demand for insights into information manipulation, international partners expressed interest in this service. Taiwan AI Labs further developed its capabilities into the Infodemic platform, providing real-time and comprehensive understanding of both domestic and international information manipulation for non-technical partners. This aids in developing digital literacy and response strategies. In recent years, Taiwan AI Labs has continued to use the Infodemic platform to observe coordinated behaviors on major Taiwanese social platforms such as Facebook, YouTube, X (Twitter), TikTok, and PTT. It employs LLMs to comprehend the targets and patterns of information manipulation attacks the responses of mainstream media. It timely records the battlefields of information warfare participated in by troll groups, along with their potential impacts.



Graph 19: Overview of the data analysis process flow on the Infodemic platform.

- This report used data and tools in <https://infodemic.cc>
- How does the system work <https://infodemic.cc/en/faq>
- DISARM Disinformation Analysis and Risk Management is an open-source framework designed for describing and understanding the behavior parts of FIMI/disinformation. It sets out best practices for fighting disinformation through sharing data & analysis, and can inform effective action. The Framework has been developed, drawing on global cybersecurity best practices. <https://www.disarm.foundation/>