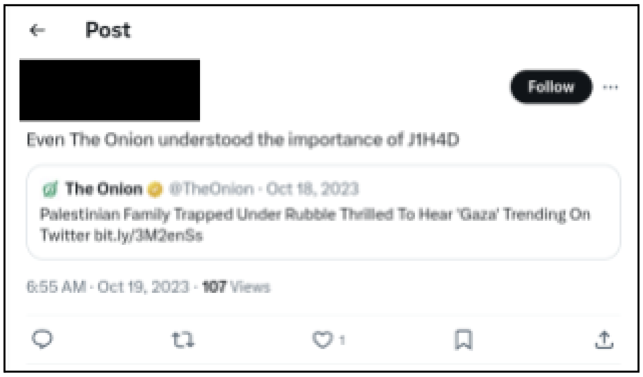
# Nisos: Top 5 Social Media Content Evasion Tactics

1 February 2024

Users spreading violative content rely on various tactics to evade content moderation on social media platforms, including to spread violent rhetoric and malign influence narratives. Users are quick to adapt to Trust and Safety teams' efforts and find new strategies to keep content that violates community guidelines on platforms. Nisos researchers monitor user messaging and routinely observe actors deploying the following tactics:

### 1. Intentionally Misspell and Use Leetspeak To Convey Blocked Terms

Trust and Safety teams may monitor for certain terms linked to content that violates community guidelines. This may include content promoting harmful activities and messaging. Some examples include rendering "Hitler" as "Jitler," "jihad" as "J1H4D," or "hoax" as "h0ax" to avoid detection.



← Post

[redacted]

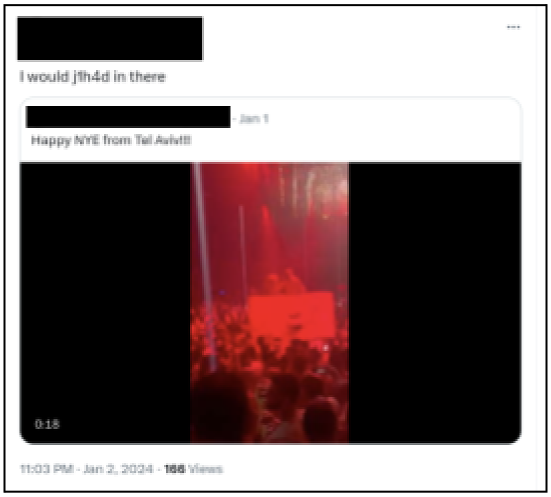
Follow ···

Even The Onion understood the importance of J1H4D

The Onion @TheOnion · Oct 18, 2023
Palestinian Family Trapped Under Rubble Thrilled To Hear 'Gaza' Trending On Twitter bit.ly/3M2enSs

6:55 AM · Oct 19, 2023 · 107 Views

1



[redacted] ···

I would j1h4d in there

[redacted] · Jan 1
Happy NYE from Tel Aviv!!!

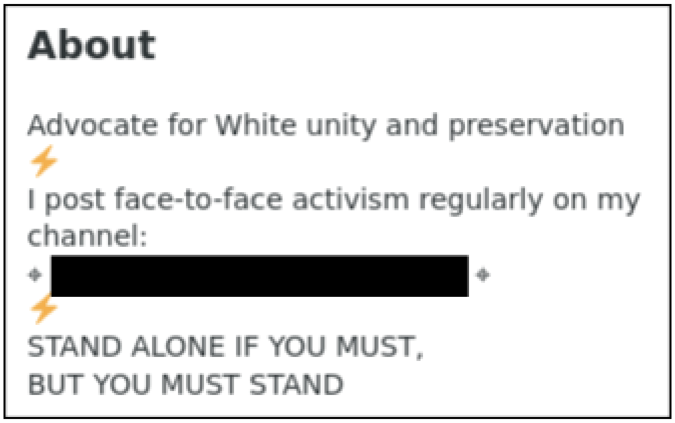
0:18

11:03 PM · Jan 2, 2024 · 166 Views

***Graphics 1 and 2: Examples of users rendering the term "jihad" using a combination of letters and numbers.***

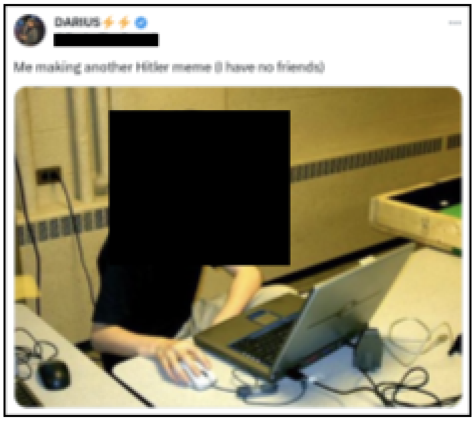
### 2. Use Coded Language

Symbols convey meaning and frequently signal membership in a group or adherence to an ideology. Actors leverage emojis to express more coded messaging understood by members of the in-group, such as "⚡⚡" as a neo-Nazi reference to the "SS" or "🕸" as a symbol for a neo-Nazi black sun. The black flag emoji, "🏴," may also signify Islamic terrorism affiliations in some cases.



About

Advocate for White unity and preservation
⚡
I post face-to-face activism regularly on my channel:
[redacted]
⚡
STAND ALONE IF YOU MUST,
BUT YOU MUST STAND



DARIUS ⚡⚡
[redacted]

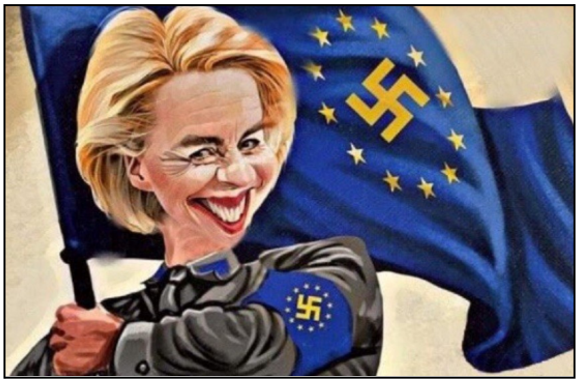
Me making another Hitler meme (I have no friends)

[redacted]

***Graphics 3 and 4: Users including lightning bolt emojis, common across white nationalist and neo-Nazi messaging.***

### 3. Post Memes To Amplify Malign Influence Narratives

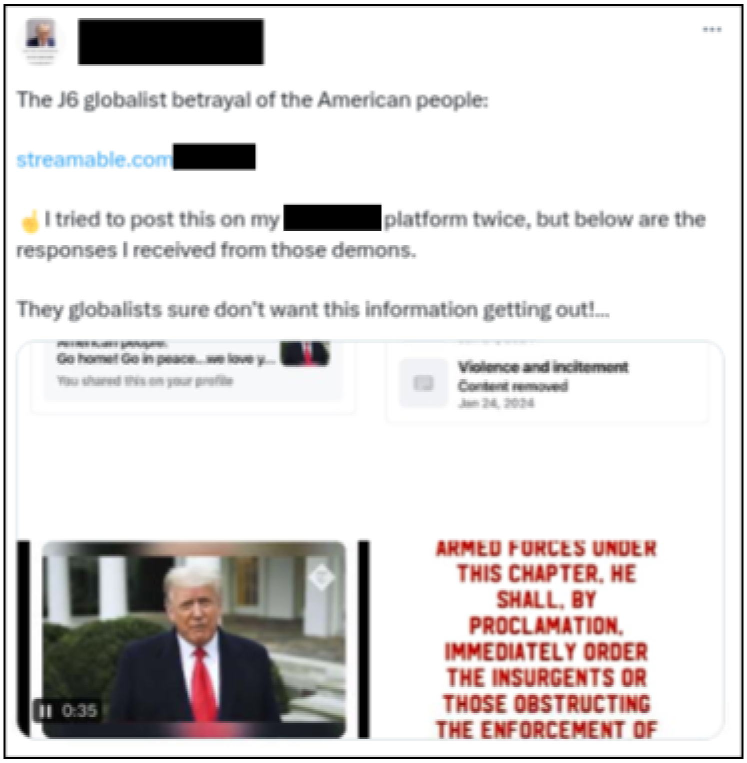
Under the guise of humor, users post memes or label content as a "joke" to convey harmful messaging and even amplify malign influence narratives. Groups of users posting the same meme and exhibiting other elements consistent with coordinated inauthentic behavior may also indicate a network boosting a common theme.



***Graphic 5: Image used with messaging discrediting European Commission President Ursula von der Leyen disseminated across social media platforms and domains.***

### 4. Link To Off-Platform Channels and Domains

Actors link to domains and communication channels on other platforms, including in the comments section, instead of only posting original content. Actors may comment on their own post with additional information and links, or they may include such details in the comments section on another user's post. Examples include links to Telegram channels, where more targeted recruitment efforts may occur, or comments with contact details to purchase illicit goods on another user's posted content.



[redacted]

The J6 globalist betrayal of the American people:

streamable.com[redacted]

☝ I tried to post this on my [redacted] platform twice, but below are the responses I received from those demons.

They globalists sure don't want this information getting out!...

American people:
Go home! Go in peace...we love y...
You shared this on your profile

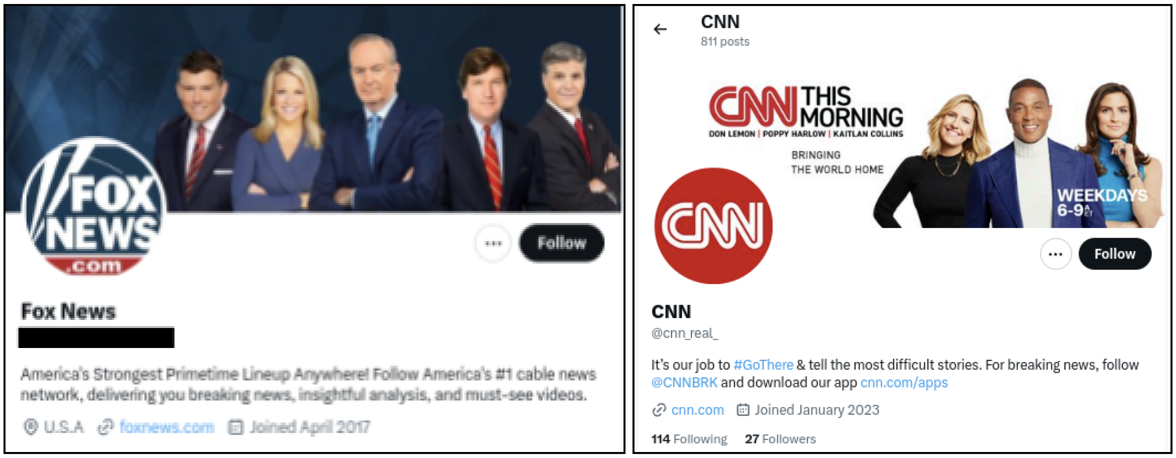
Violence and incitement
Content removed
Jan 24, 2024

ARMED FORCES UNDER THIS CHAPTER, HE SHALL, BY PROCLAMATION, IMMEDIATELY ORDER THE INSURGENTS OR THOSE OBSTRUCTING THE ENFORCEMENT OF

***Graphic 6: User shares an off-platform link to content that had been removed for violence and incitement.***

### 5. Impersonate Public Figures and Known Institutions To Appear Credible

Actors create accounts impersonating political leaders, public figures, media outlets, and institutions often to gain credibility among readers and viewers. Actors may also leverage these accounts to conceal their identities and challenge attribution efforts. For the same reason, actors may use rented or temporary accounts to amplify specific narratives. Actors operating these accounts may pay for advertising to target a specific demographic or purchase bot engagements to broaden the content's reach.



Fox News
[redacted]

America's Strongest Primetime Lineup Anywhere! Follow America's #1 cable news network, delivering you breaking news, insightful analysis, and must-see videos.

U.S.A foxnews.com Joined April 2017

CNN
811 posts

CNN
@cnn_real_

It's our job to #GoThere & tell the most difficult stories. For breaking news, follow @CNNBRK and download our app cnn.com/apps

cnn.com Joined January 2023

114 Following 27 Followers

***Graphics 7 and 8: Examples of inauthentic accounts impersonating well-known news networks.***

Users are adaptable and deploy various techniques to post and amplify violative and malign influence content across social media. It is essential to remain aware of the evolutions in user behavior aimed at evading Trust and Safety protections. Nisos provides services to support Trust and Safety teams to stay abreast of threats to their platform.