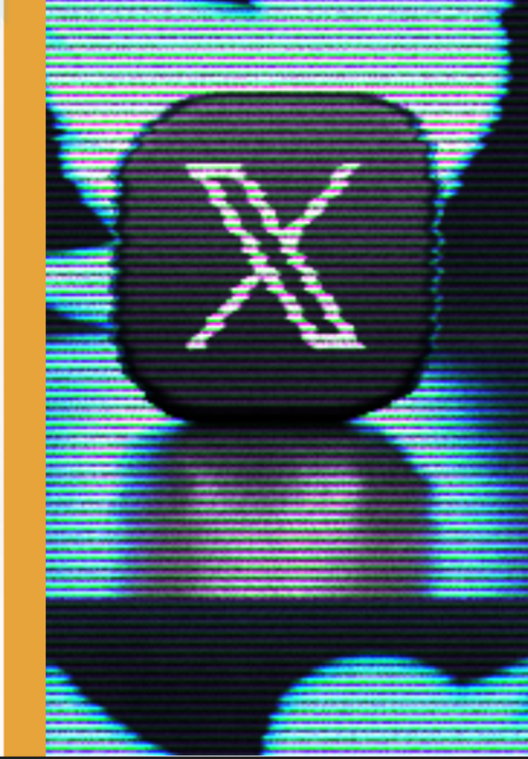
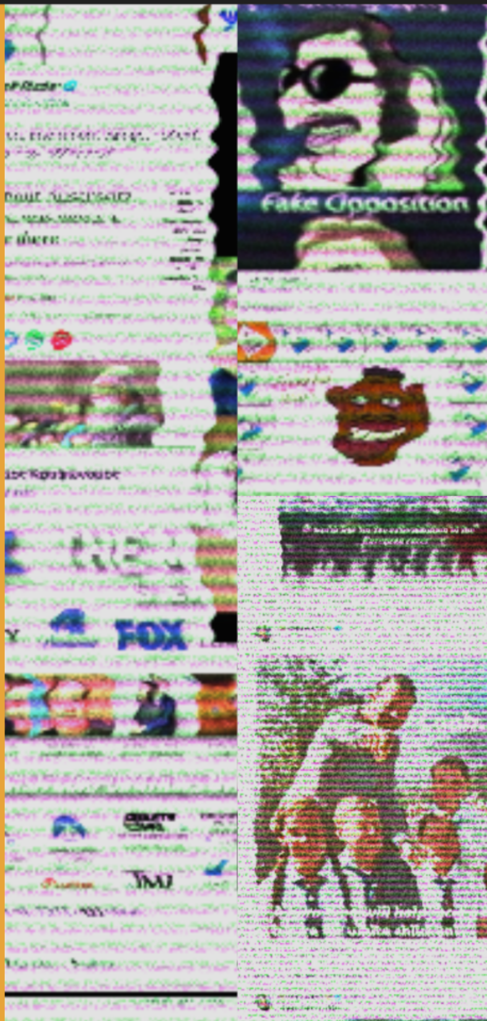
Center for Countering Digital Hate



X CONTENT MODERATION FAILURE

How Twitter/X continues to host posts reported for extreme hate speech



The Center for Countering Digital Hate works to stop the spread of online hate and disinformation through innovative research, public campaigns and policy advocacy.

Our mission is to protect human rights and civil liberties online.

Social media platforms have changed the way we communicate, build and maintain relationships, set social standards, and negotiate and assert our society's values. In the process, they have become safe spaces for the spread of hate, conspiracy theories and disinformation.

Social media companies erode basic human rights and civil liberties by enabling the spread of online hate and disinformation.

At CCDH, we have developed a deep understanding of the online harm landscape, showing how easily hate actors and disinformation spreaders exploit the digital platforms and search engines that promote and profit from their content.

We are fighting for better online spaces that promote truth, democracy, and are safe for all. Our goal is to increase the economic and reputational costs for the platforms that facilitate the spread of hate and disinformation.

If you appreciate this report, you can donate to CCDH at counterhate.com/donate. In the United States, Center for Countering Digital Hate Inc is a 501(c)(3) charity. In the United Kingdom, Center for Countering Digital Hate Ltd is a non-profit company limited by guarantee.

Published September 2023
© Center for Countering Digital Hate Inc

Please Read

This report includes content on the following themes which may be distressing to readers:

- Violence
- Antisemitism
- Holocaust denial
- Racism

Reader discretion is advised.

Contents

X continues to host 86% of 300 posts reported for extreme hate speech

Research has found that X, formerly Twitter, continues to host 86% of 300 posts reported for hate speech, despite a commitment to combat content “motivated by hatred, prejudice or intolerance”.

Researchers collected a total of 300 posts promoting hate from 100 X accounts, amounting to three posts per account. These posts were categorized as promoting antisemitism, anti-black hatred, other racism, neo-Nazism and white supremacism. Together the accounts identified have a sum total of 1,060,106 followers.

Posts were independently assessed by two researchers to determine whether they represent a violation of X’s policies on hate speech, with only those that both agreed were violations included in the final set of 300 posts. Posts were reported to X using its user reporting tools on August 30 and 31, and reviewed for action taken against them one week later on September 7.

CCDH has provided a full database of the posts it selected, and has taken screenshots. They will be provided, on request, to any journalist or researcher enquiring.

One week after reporting, researchers found that X had continued to host 86.33% (259) of the 300 posts and 90.00% (90) of the 100 accounts remained active. X continued to host posts that were:

- Promoting racist caricatures of black people and Jewish people
- Denying the Holocaust or mocking victims of the Holocaust
- Labeling Hitler as “A hero who will help secure a future for white children!”
- Claiming “Blacks don’t need provoking before becoming violent. It’s in their nature.”
- Encouraging users to “Stop Race mixing” and “break up with your non-white gf today”
- Promoting conspiracies that Jews promote mass migration and “control the blacks”
- Claiming Jews exert “victimhood and cult-like behavior” and suffer from “schizophrenia”
- Memes accusing Black people of being harmful to “A quiet functioning society”

X continued to host these posts even after they were reported, despite them clearly violating the platform’s policies against hateful content, which prohibit racist slurs, dehumanization, and hateful imagery such as the Nazi swastika.

Researchers received notifications from X that three accounts reported as part of the study had been "locked", stating that "they can't post, repost, or Like content, and we'll ask them to remove the reported content if they want to regain full access to their account." As of September 7, this has resulted in posts that promote Holocaust denial, including one that mocks Anne Frank's status as a victim of the Holocaust, remaining publicly visible on the platform.

Musk recently posted: "I'm pro free speech, but against anti-Semitism of any kind", whilst X's new CEO, Linda Yaccarino, claimed that X is "a much healthier and safer platform than it was a year ago". However, previous research from the CCDH found that the volume of posts containing hate speech surged after Elon Musk's acquisition of Twitter.

X continues to host antisemitic content including Holocaust denial

Researchers identified a total of 140 posts that promoted antisemitism, including racist caricatures of Jewish people and claims that Jews control the world. X continued hosting this content in 85.00% (119) of cases.

X continued to host 16 posts that researchers identified as containing Holocaust denial, including posts that mock victims of Holocaust and present death camps as benign.

This comes as X's owner Elon Musk declared his intention to sue the Anti-Defamation League (ADL), America's oldest organization working to counter antisemitism, over their criticism of X for allowing a boom in anti-Jewish hate on the platform.

Researchers identify 38 ads placed next to hateful posts

CCDH researchers also identified 38 ads from companies, such as Apple and Disney, that were placed adjacent to hateful posts. The ads appeared either on the "For You" feed or on the profiles of accounts promoting hate speech.

On the point of brand safety on X, Yaccarino said: "since acquisition, we [X] have built brand safety and content moderation tools that have never existed before at this company."

In hopes of winning back advertisers, X recently signed a new brand safety deal with digital ad-tech firm Integral Ad Science and introduced an industry-standard blocklist that claims to "protect advertisers from appearing adjacent to unsafe keywords in the Home Timeline".

Methodology

- The full set of posts used in this analysis is available below. **Note that content includes extreme antisemitism, anti-black hatred, white supremacism and neo-Nazism.**

  X Content Moderation dataset, CCDH, 7 September 2023, https://docs.google.com/spreadsheets/d/1Uu4yh9YeeYN2uWEBsqwohf33D9Y2JikfnrO3_InrRYc/edit?usp=sharing

- All 300 posts, X accounts and ads included in this analysis can be found in the link above. Posts were reported to X on August 30th and 31st, and reviewed to see what action X had taken at least one week later on September 7, examining both notifications from X on action against reports, and relevant posts or accounts themselves to examine whether they had been removed or otherwise acted upon.

- For each post, researchers recorded whether the following actions had been taken:
  - No action
  - Actioned, noting the action taken
    - Post removed
    - Account suspended
    - Account locked (see below for further explanation)

- Where it was unclear whether a post or account had been removed by X or the user controlling the account, this was counted as "actioned" to ensure that all possible actions taken by X were accounted for.

- Researchers identified hateful accounts and content by examining the content being shared by known hateful accounts and account recommendations presented by X on hateful profiles under the "Who to follow" and "You might like" panels.

- Each post was categorized according to the types of hate it promotes, and then checked by a second researcher to ensure it represented a clear violation of X's policies on hate speech (see below). Researchers stopped collecting posts when they amassed 300 posts from 100 unique accounts that met these criteria.

- X continued to host posts identified by researchers that clearly violate the platform's policies against hateful conduct, which prohibits:
  - "[attacks on] other people on the basis of race, ethnicity, national origin"
  - "dehumanization" on the basis of race, religion, nationality and other qualities

    - Hateful imagery such as the Nazi swastika

- The policy also contains a commitment to “combating abuse motivated by hatred, prejudice or intolerance, particularly abuse that seeks to silence the voices of those who have been historically marginalized.”

  Hateful Conduct, X, April 2023, *https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy*

Background

- The new CEO of X, Linda Yaccarino, recently claimed in an interview that X is "a much healthier and safer platform than it was a year ago".

  "X Corp. now a much healthier and safe platform than a year ago says Lina Yaccarino", CNBC, 10 August 2023, 01:11, https://www.cnbc.com/video/2023/08/10/x-corp-now-a-much-healthier-and-safer-platform-than-a-year-ago-says-linda-yaccarino.html

- However, previous research from the CCDH found that the volume of posts containing hate speech surged after Elon Musk's acquisition of Twitter.

  The Musk Bump: Quantifying the rise in hate speech under Elon Musk, CCDH, 6 December 2022, https://counterhate.com/blog/the-musk-bump-quantifying-the-rise-in-hate-speech-under-elon-musk/

- To brand safety on X, Yaccarino said that "since acquisition, we [X] have built brand safety and content moderation tools that have never existed before at this company."

  "X Corp. now a much healthier and safe platform than a year ago says Lina Yaccarino", CNBC, 10 August 2023, 01:18, https://www.cnbc.com/video/2023/08/10/x-corp-now-a-much-healthier-and-safer-platform-than-a-year-ago-says-linda-yaccarino.html

- In hopes of winning back advertisers, X recently signed a new brand safety deal with digital ad-tech firm Integral Ad Science and introduced an industry-standard blocklist that alleges to "protect advertisers from appearing adjacent to unsafe keywords in the Home Timeline".

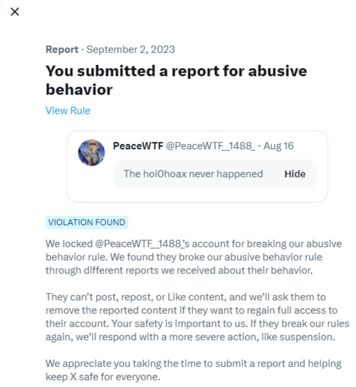
  A safer X is a better X, X, accessed 1 September 2023, https://business.twitter.com/en/help/ads-policies/brand-safety.html

- X's owner Elon Musk has declared his intention to sue the Anti-Defamation League (ADL), America's oldest organization working to counter antisemitism, over their criticism of X for allowing a boom in anti-Jewish hate on the platform.

  "Elon Musk blames the ADL for 60% ad sales decline at X, threatens to sue", CNN, 5 September 2023, https://www.cnn.com/2023/09/05/tech/elon-musk-adl-lawsuit/index.html

X’s decision to “lock” accounts in response to reports

- Researchers received notifications from X that three accounts reported as part of the study had been “locked”, stating that “they can’t post, repost, or Like content, and we’ll ask them to remove the reported content if they want to regain full access to their account.” Note that the date in screenshots is the date the notification was received.



Report · September 2, 2023

You submitted a report for abusive behavior

View Rule

PeaceWTF @PeaceWTF_1488_ · Aug 16

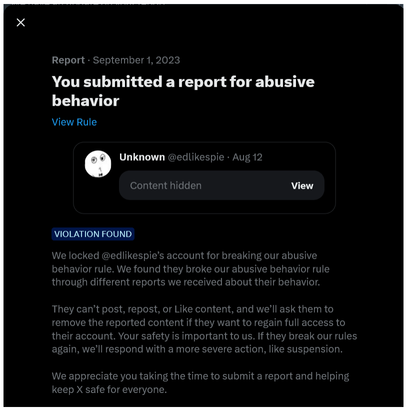
The hol0hoax never happened **Hide**

VIOLATION FOUND

We locked @PeaceWTF_1488_’s account for breaking our abusive behavior rule. We found they broke our abusive behavior rule through different reports we received about their behavior.

They can’t post, repost, or Like content, and we’ll ask them to remove the reported content if they want to regain full access to their account. Your safety is important to us. If they break our rules again, we’ll respond with a more severe action, like suspension.

We appreciate you taking the time to submit a report and helping keep X safe for everyone.



Report · September 1, 2023

You submitted a report for abusive behavior

View Rule

Unknown @edlikespie · Aug 12

Content hidden **View**

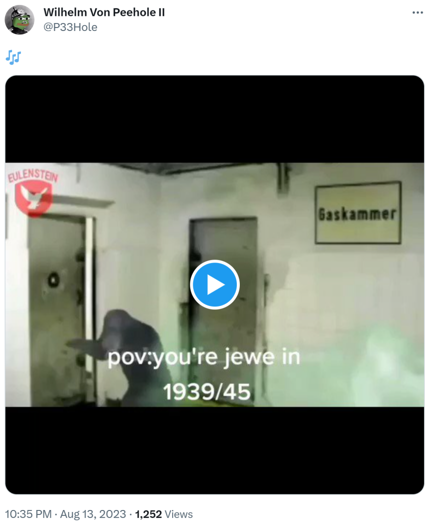
VIOLATION FOUND

We locked @edlikespie’s account for breaking our abusive behavior rule. We found they broke our abusive behavior rule through different reports we received about their behavior.

They can’t post, repost, or Like content, and we’ll ask them to remove the reported content if they want to regain full access to their account. Your safety is important to us. If they break our rules again, we’ll respond with a more severe action, like suspension.

We appreciate you taking the time to submit a report and helping keep X safe for everyone.

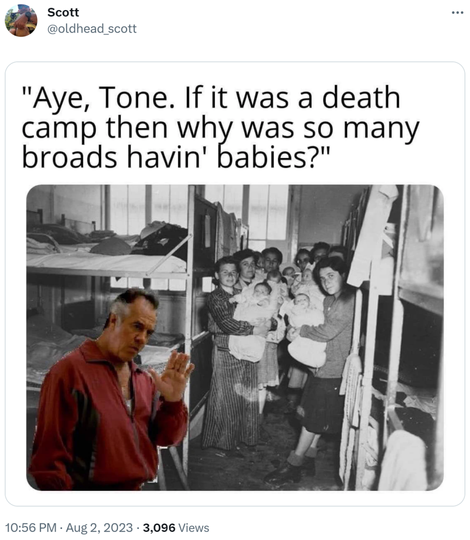
- As of September 7, this has resulted in posts that promote Holocaust denial and mock victims of the Holocaust remaining publicly visible on the platform.



Wilhelm Von Peehole II
@P33Hole

🎶

10:35 PM · Aug 13, 2023 · **1,252** Views



Scott
@oldhead_scott

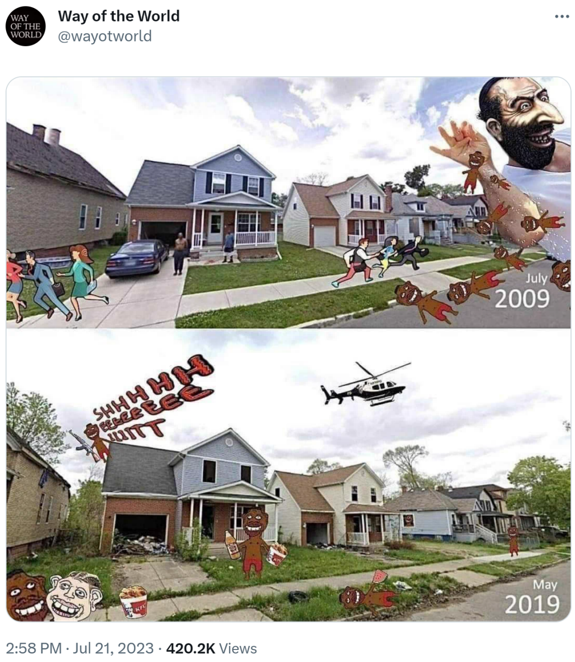
"Aye, Tone. If it was a death camp then why was so many broads havin' babies?"

10:56 PM · Aug 2, 2023 · **3,096** Views

@P33Hole, X, 14 August 2023,
https://drive.google.com/file/d/1EuksNb39yr-J3Mbr6UcENZ9tvsmnOeib/view
@oldhead_scott, X, 2 August 2023,
https://drive.google.com/open?id=1g7pJ27dkmogS9CiQXGVdo1CFmxempaWj

Examples of hateful posts still hosted by X as of September 7

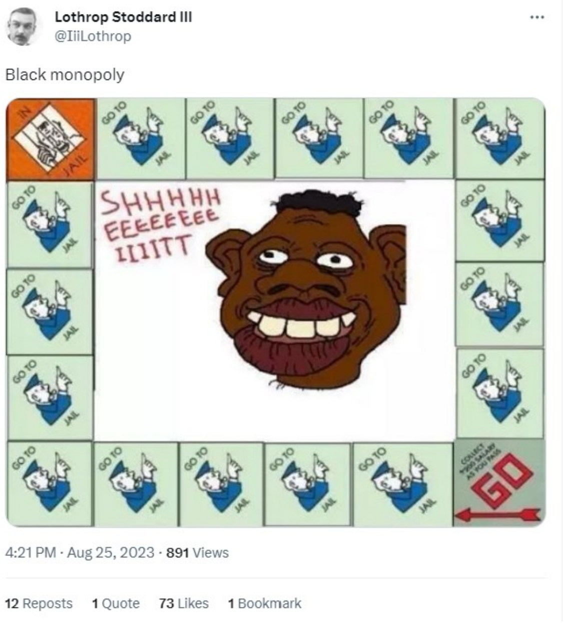
- Promoting racist caricatures of black people and Jewish people.



Way of the World
@wayotworld

2:58 PM · Jul 21, 2023 · **420.2K** Views

@wayotworld, X, 21 July 2023, *https://drive.google.com/open?id=14i6BYE9Vs4Sbp3O2jkciQclHDzeYCCA7*



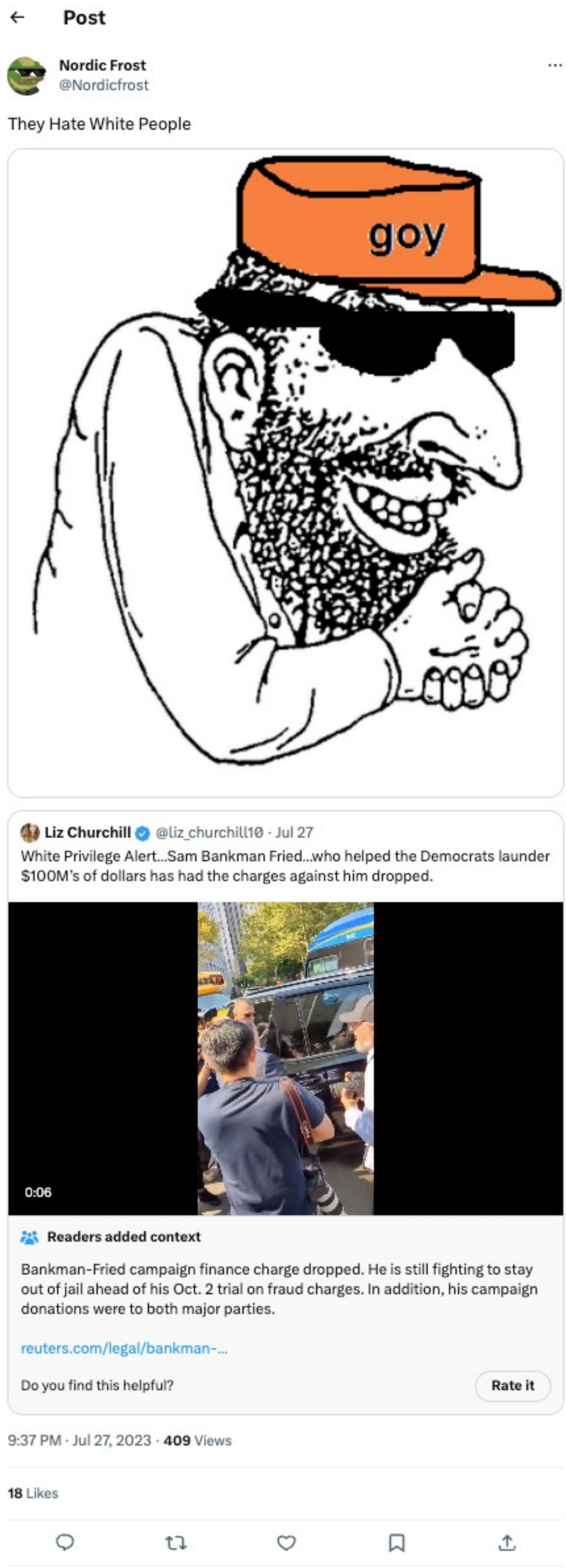
Lothrop Stoddard III
@IiiLothrop

Black monopoly

4:21 PM · Aug 25, 2023 · **891** Views

12 Reposts **1** Quote **73** Likes **1** Bookmark

IiiLothrop, X, 25 August 2023, *https://drive.google.com/file/d/1oysZhmdjFGNYJjjF-HEtOecJqVpUoROk/view*



← Post

Nordic Frost
@Nordicfrost

They Hate White People

Liz Churchill @liz_churchill10 · Jul 27
White Privilege Alert...Sam Bankman Fried...who helped the Democrats launder $100M's of dollars has had the charges against him dropped.

Readers added context

Bankman-Fried campaign finance charge dropped. He is still fighting to stay out of jail ahead of his Oct. 2 trial on fraud charges. In addition, his campaign donations were to both major parties.

reuters.com/legal/bankman-...

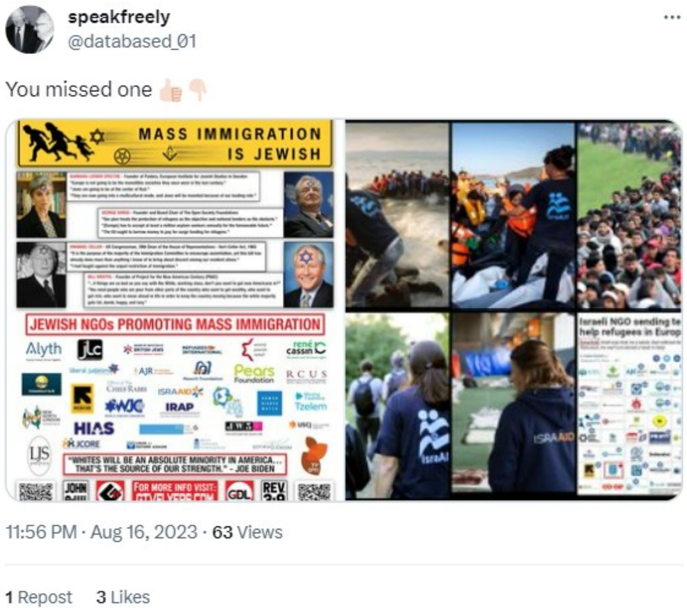
Do you find this helpful?

9:37 PM · Jul 27, 2023 · **409** Views

18 Likes

Nordicfrost, X, 27 July 2023,
https://drive.google.com/file/d/1dFrDfbBG9271CQbDdhqe0bgSOwUiQv64/view

- Promoting conspiracies that Jews promote mass migration and “control the blacks”



speakfreely
@databased_01

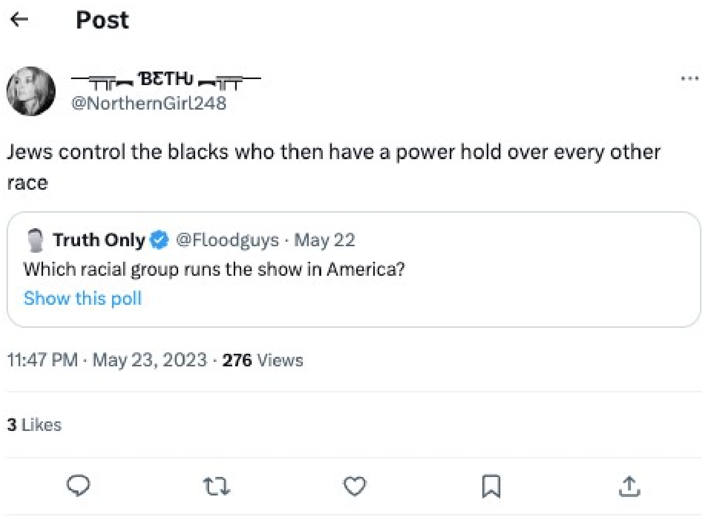
You missed one 👍👎

MASS IMMIGRATION IS JEWISH
JEWISH NGOs PROMOTING MASS IMMIGRATION
Alyth
jlc
Pears Foundation
RCUS
AJR
ISRAAID
#WJC
IRAP
Tzelem
HIAS
JCORE
LJS
"WHITES WILL BE AN ABSOLUTE MINORITY IN AMERICA... THAT'S THE SOURCE OF OUR STRENGTH." - JOE BIDEN
FOR MORE INFO VISIT:
GDL
REV
Israeli NGO sending te help refugees in Europ
IsraAID
ISRAAID

11:56 PM · Aug 16, 2023 · 63 Views

1 Repost 3 Likes

Databased_01, X, 16 August 2023, [*https://drive.google.com/file/d/1b9EtaNGUdhDxsnZ2AIR4XZo-Tx1XOcPw/view*](https://drive.google.com/file/d/1b9EtaNGUdhDxsnZ2AIR4XZo-Tx1XOcPw/view)



← Post

—╥╥━ 'BETH' ━╥╥—
@NorthernGirl248

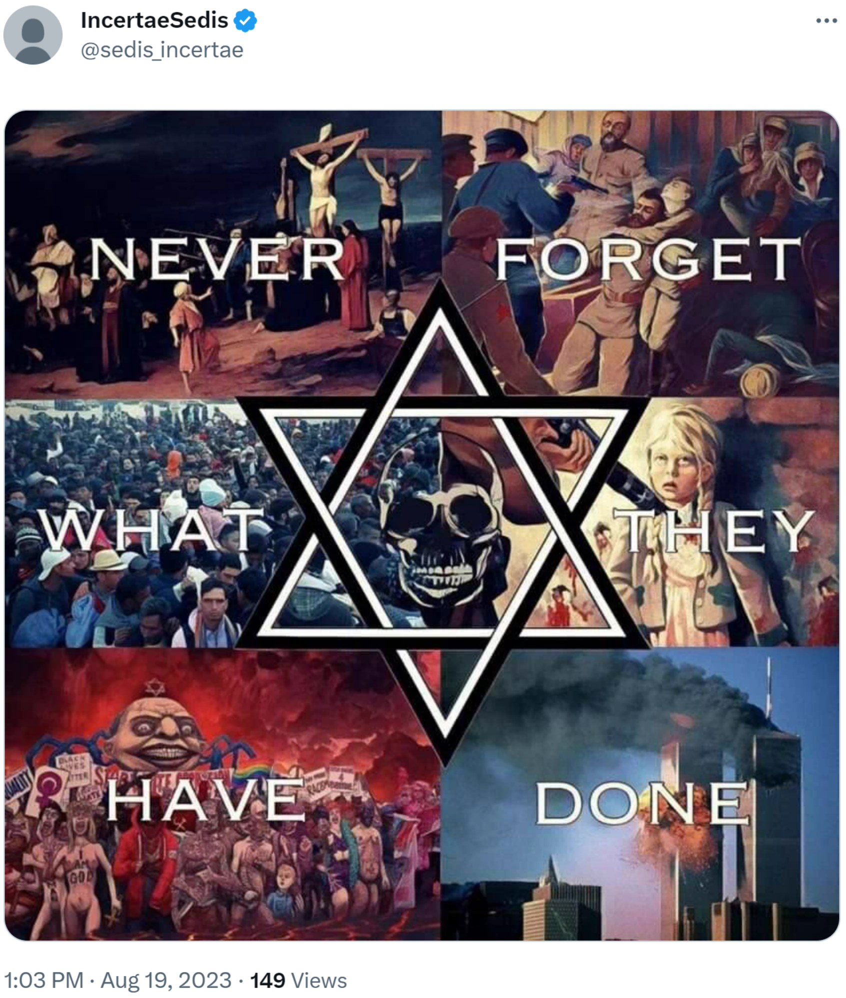
Jews control the blacks who then have a power hold over every other race

Truth Only @Floodguys · May 22
Which racial group runs the show in America?
Show this poll

11:47 PM · May 23, 2023 · 276 Views

3 Likes

NorthernGirl248, X, 23 May 2023, [*https://drive.google.com/file/d/1oKNgv4hJLkVccbiYHcDFfmvrvt3dXC4T/view*](https://drive.google.com/file/d/1oKNgv4hJLkVccbiYHcDFfmvrvt3dXC4T/view)

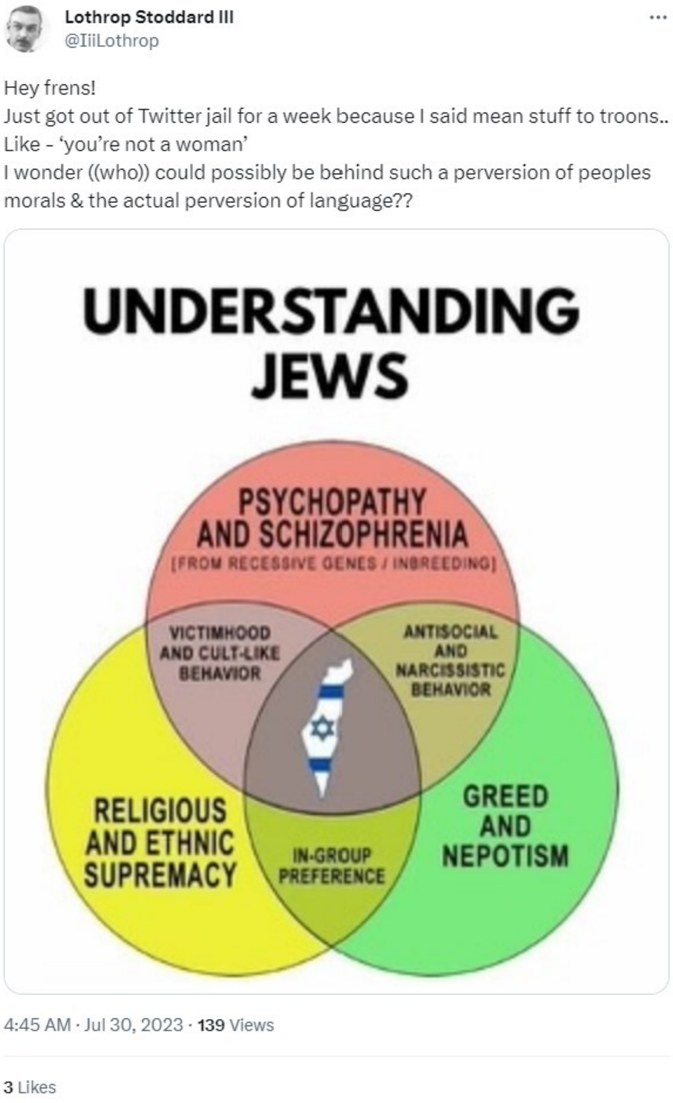


IncertaeSedis
@sedis_incertae

1:03 PM · Aug 19, 2023 · **149** Views

@sedis_incertae, X, 19 August 2023,
https://drive.google.com/open?id=1CBSQmcOCM8V67DGXNZI8OgjuRudLXMhN

- Claiming Jews exert “victimhood and cult-like behavior” and suffer from “Schizophrenia”



Lothrop Stoddard III
@IiiLothrop

Hey frens!
Just got out of Twitter jail for a week because I said mean stuff to troons..
Like - ‘you’re not a woman’
I wonder ((who)) could possibly be behind such a perversion of peoples morals & the actual perversion of language??

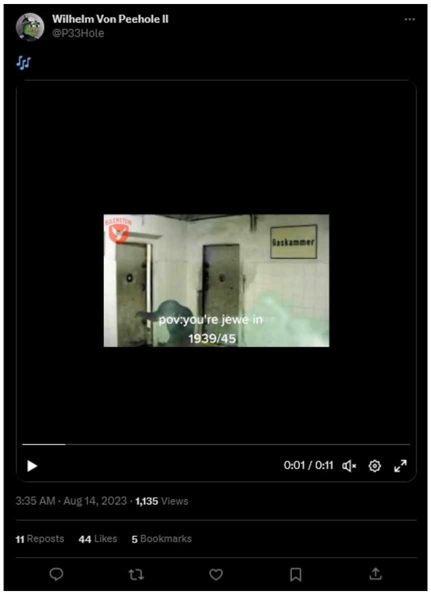
UNDERSTANDING JEWS
PSYCHOPATHY AND SCHIZOPHRENIA
(FROM RECESSIVE GENES / INBREEDING)
VICTIMHOOD AND CULT-LIKE BEHAVIOR
ANTISOCIAL AND NARCISSISTIC BEHAVIOR
RELIGIOUS AND ETHNIC SUPREMACY
IN-GROUP PREFERENCE
GREED AND NEPOTISM

4:45 AM · Jul 30, 2023 · **139** Views

3 Likes

IiiLothrop, X, 30 July 2023, *https://drive.google.com/file/d/1KRsu1-5WEgQKog7LEc3YRmpOLKO5YAQ4/view*

- Denying the Holocaust or mocking victims of the Holocaust



Wilhelm Von Peehole II
@P33Hole

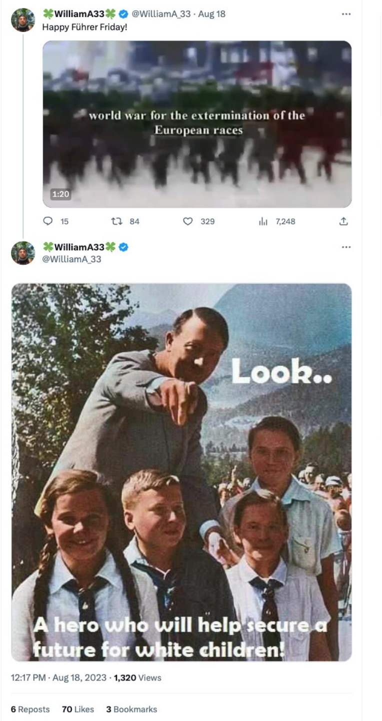
🎶

3:35 AM · Aug 14, 2023 · 1,135 Views

11 Reposts 44 Likes 5 Bookmarks

P33Hole, X, 14 August 2023,
https://drive.google.com/file/d/1EuksNb39yr-J3Mbr6UcENZ9tvsmnOeib/view

- Labeling Hitler as “A hero who will help secure a future for white children!”



☘WilliamA33☘ @WilliamA_33 · Aug 18
Happy Führer Friday!

world war for the extermination of the European races
1:20

15 84 329 7,248

☘WilliamA33☘
@WilliamA_33

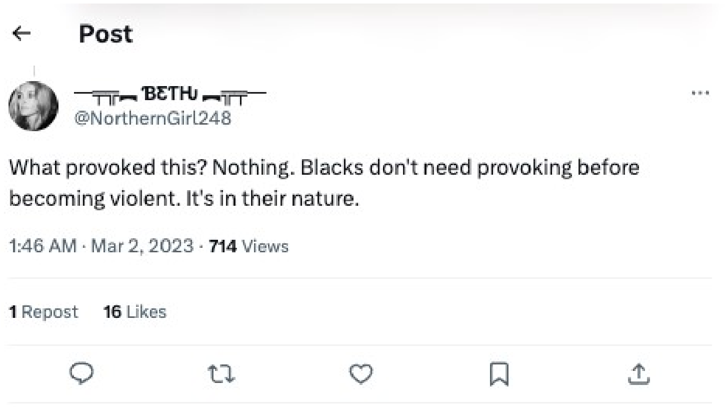
Look..
A hero who will help secure a future for white children!

12:17 PM · Aug 18, 2023 · 1,320 Views

6 Reposts 70 Likes 3 Bookmarks

WilliamA_33, X, 18 August 2023,
https://drive.google.com/file/d/1rHifb5djnJP-JHtg84yHwaNUVStB85so/view

- Claiming “Blacks don’t need provoking before becoming violent. It’s in their nature.”



Post

BETH

@NorthernGirl248

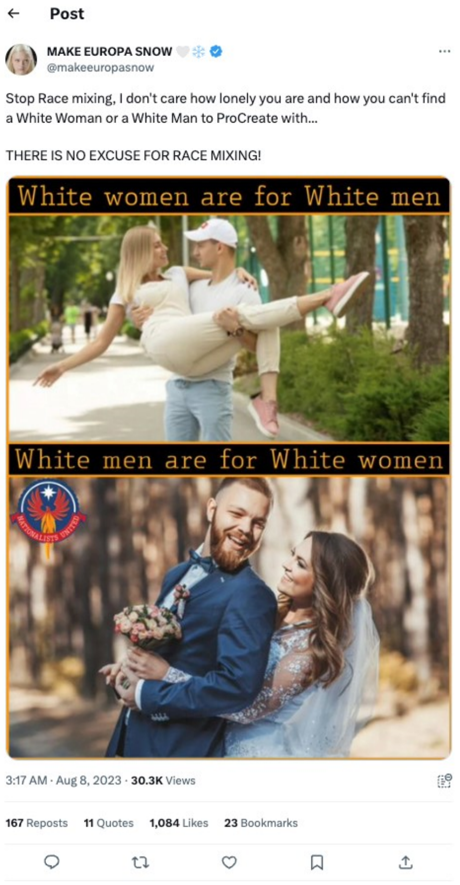
What provoked this? Nothing. Blacks don't need provoking before becoming violent. It's in their nature.

1:46 AM · Mar 2, 2023 · **714** Views

1 Repost **16** Likes

NorthernGirl248, X, 2 March 2023, [*https://drive.google.com/file/d/1OgPoSyaClPXnY4LYVt3FsdDh_fKgNPBo/view*](https://drive.google.com/file/d/1OgPoSyaClPXnY4LYVt3FsdDh_fKgNPBo/view)

- Encouraging users to “Stop Race mixing” and “break up with your non-white gf today”



Post

MAKE EUROPA SNOW

@makeeuropasnow

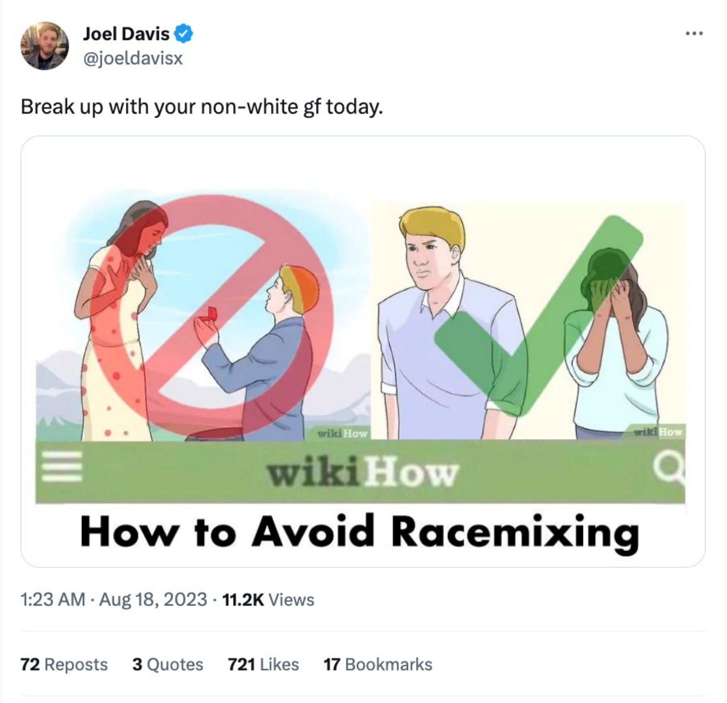
Stop Race mixing, I don't care how lonely you are and how you can't find a White Woman or a White Man to ProCreate with...

THERE IS NO EXCUSE FOR RACE MIXING!

3:17 AM · Aug 8, 2023 · **30.3K** Views

167 Reposts **11** Quotes **1,084** Likes **23** Bookmarks

Makeeuropasnow, X, 8 August 2023, [*https://drive.google.com/file/d/12jdR9XVDDVilx5Kk87wBleLpekj57zRC/view*](https://drive.google.com/file/d/12jdR9XVDDVilx5Kk87wBleLpekj57zRC/view)



Joel Davis
@joeldavisx

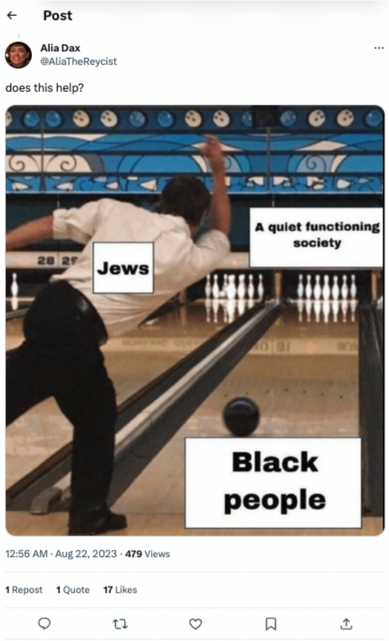
Break up with your non-white gf today.

1:23 AM · Aug 18, 2023 · **11.2K** Views

72 Reposts **3** Quotes **721** Likes **17** Bookmarks

Joeldavisx, X, 18 August 2023, [*https://drive.google.com/file/d/1-QyV8te7X-PRRUX40Ak8ZWD-XskEpkKy/view*](https://drive.google.com/file/d/1-QyV8te7X-PRRUX40Ak8ZWD-XskEpkKy/view)

- Memes accusing Black people of being harmful to "A quiet functioning society"



Post

Alia Dax
@AliaTheReycist

does this help?

12:56 AM · Aug 22, 2023 · **479** Views

1 Repost **1** Quote **17** Likes

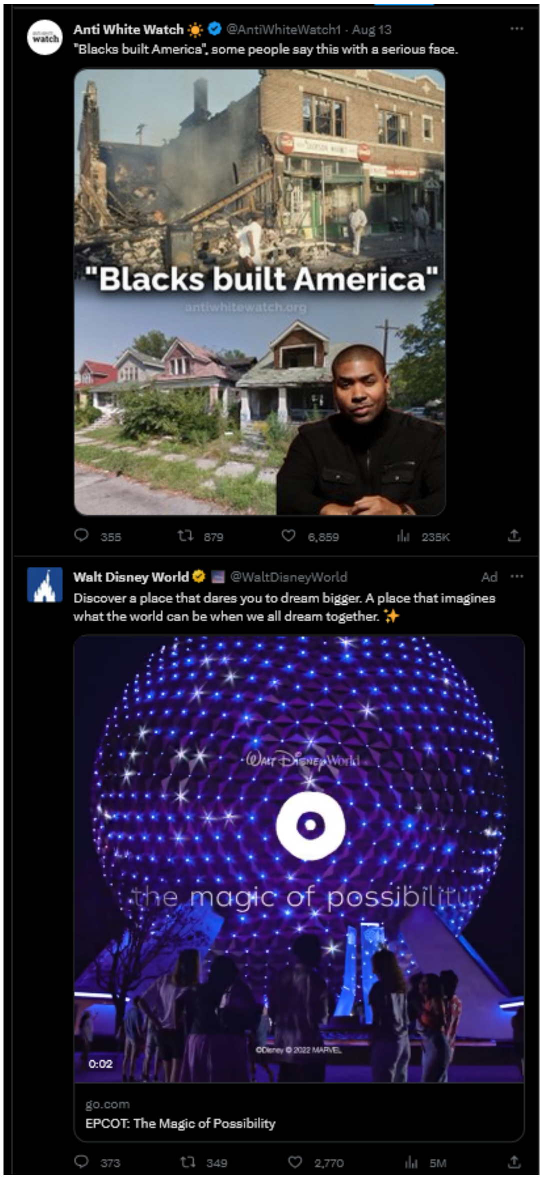
@AliaTheReycist, X, 21 August 2023, [*https://drive.google.com/open?id=1l35duyMxSa1g-aQcStOHmcf_nQVllwPM*](https://drive.google.com/open?id=1l35duyMxSa1g-aQcStOHmcf_nQVllwPM)

Examples of ads placed next to hateful posts

- The full set of ads used in this analysis is available below.

  X Content Moderation dataset, CCDH, 7 September 2023,
  https://docs.google.com/spreadsheets/d/1Uu4yh9YeeYN2uWEBsqwohf33D9Y2Jikfnr03_InrRYc/edit#gid=726502028

- An ad from Walt Disney World placed below a post mocking Black Americans



Anti White Watch ☀ @AntiWhiteWatch1 · Aug 13
"Blacks built America", some people say this with a serious face.

355 879 6,859 235K

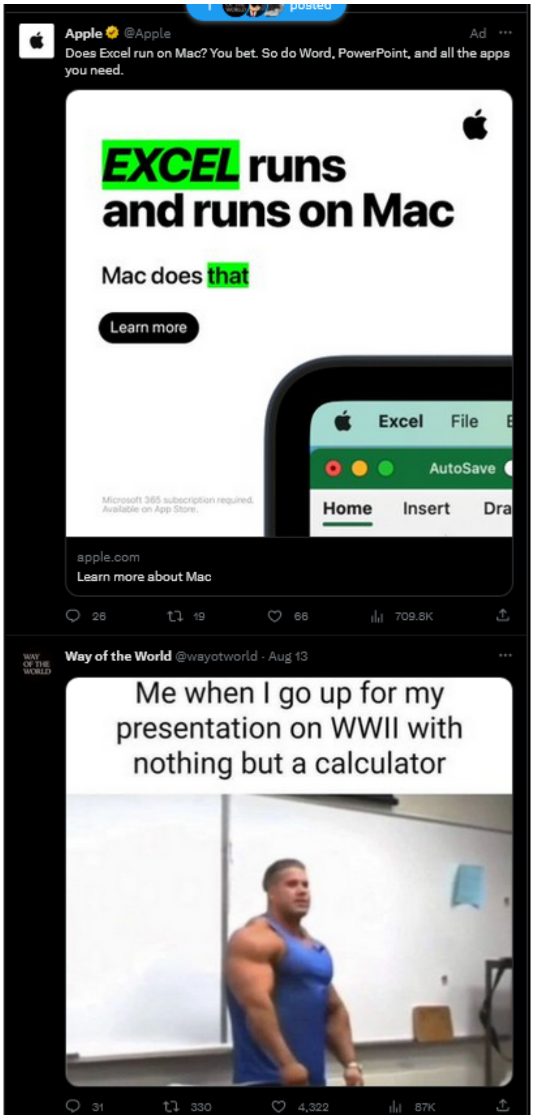
Walt Disney World @WaltDisneyWorld Ad
Discover a place that dares you to dream bigger. A place that imagines what the world can be when we all dream together. ✨

go.com
EPCOT: The Magic of Possibility

373 349 2,770 5M

AntiWhiteWatch1, X, 13 August 2023,
https://twitter.com/AntiWhiteWatch1/status/1690535047349350400
WaltDisneyWorld, X, 21 July 2023,
https://twitter.com/WaltDisneyWorld/status/1682492538018750464

- An ad from Apple placed above a post alluding to Holocaust Denial



Apple @Apple · Ad

Does Excel run on Mac? You bet. So do Word, PowerPoint, and all the apps you need.

apple.com
Learn more about Mac

26 · 19 · 66 · 709.8K

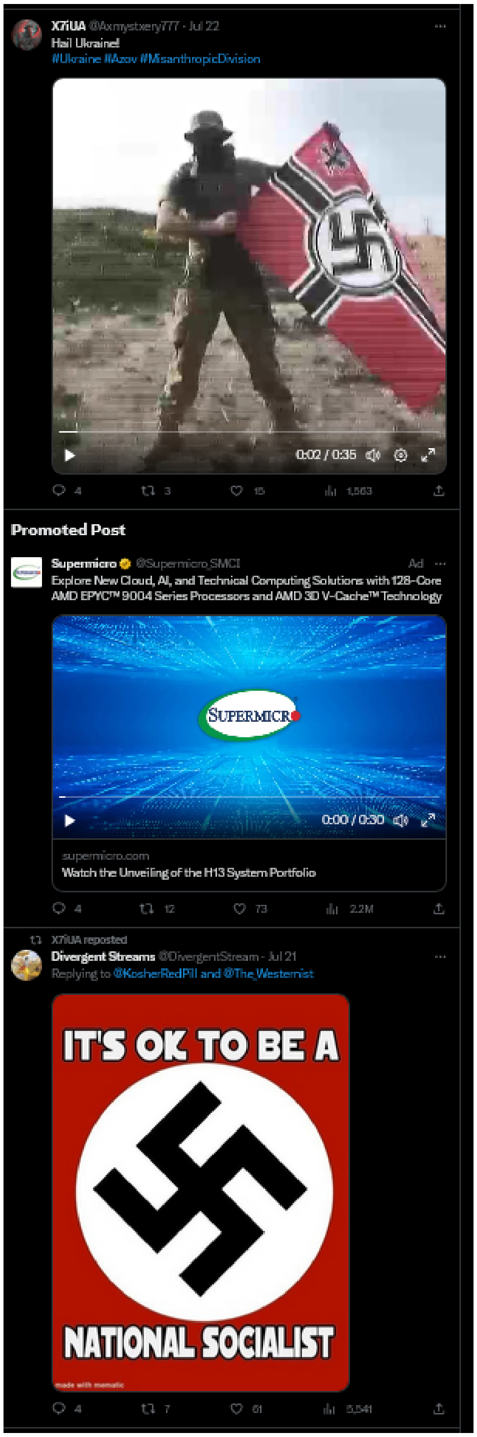
Way of the World @wayotworld · Aug 13

Me when I go up for my presentation on WWII with nothing but a calculator

31 · 330 · 4,322 · 87K

Wayotworld, X, 13 August 2023,
https://twitter.com/Apple/status/1641602026856669186
Apple, X, 31 March 2023, https://twitter.com/Apple/status/1641602026856669186

- An ad from Supermicro placed between two posts praising Nazis



X7iUA @Axmystxery777 · Jul 22

Hail Ukraine!

#Ukraine #Azov #MisanthropicDivision

Promoted Post

Supermicro @Supermicro_SMCI Ad

Explore New Cloud, AI, and Technical Computing Solutions with 128-Core AMD EPYC™ 9004 Series Processors and AMD 3D V-Cache™ Technology

supermicro.com

Watch the Unveiling of the H13 System Portfolio

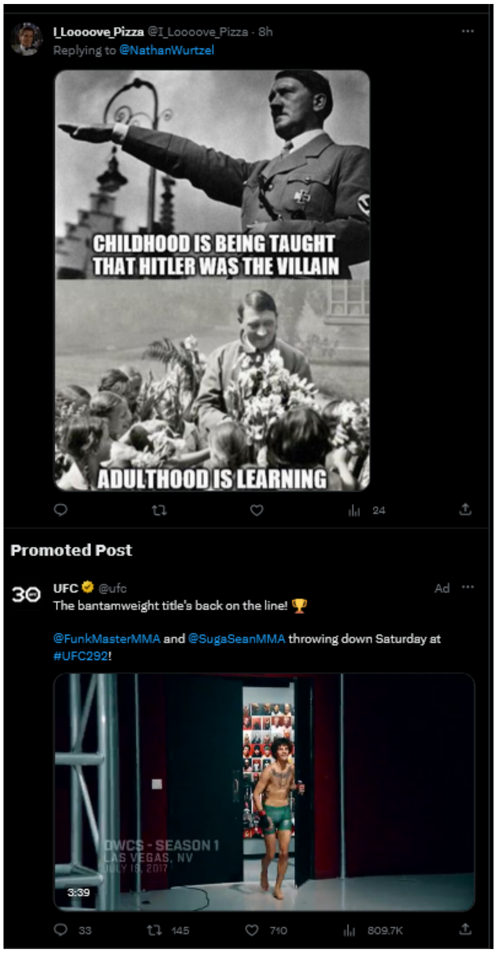
X7iUA reposted

Divergent Streams @DivergentStream · Jul 21

Replying to @KosherRedPill and @The_Westernist

Axmystxery777, X, 22 June 2023, [*https://twitter.com/Axmystxery777/status/1682719622468911104*](https://twitter.com/Axmystxery777/status/1682719622468911104)
Supermicro_SMCI, X, 4 August 2023, [*https://twitter.com/Supermicro_SMCI/status/1687552772642848771?s=20*](https://twitter.com/Supermicro_SMCI/status/1687552772642848771?s=20)

- This ad for UFC was placed next to a post praising Hitler.



I_Loooove_Pizza @I_Loooove_Pizza · 8h ···
Replying to @NathanWurtzel

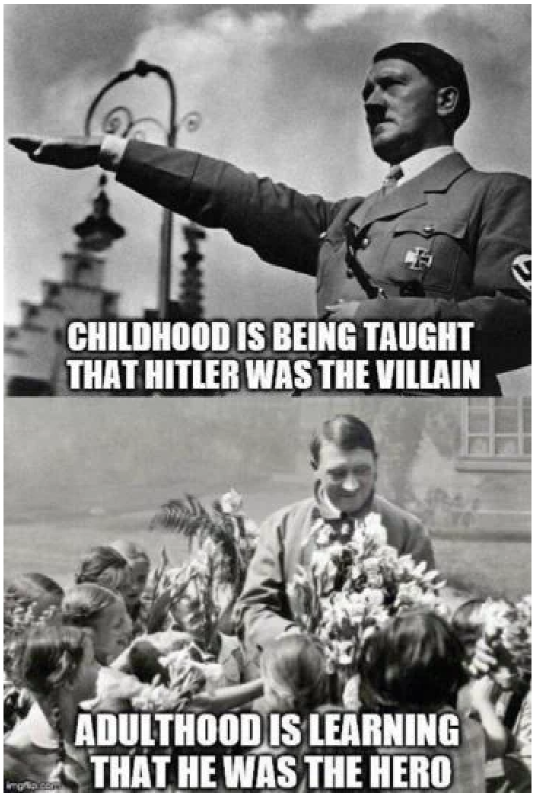
24

Promoted Post

UFC @ufc Ad ···
The bantamweight title's back on the line! 🏆

@FunkMasterMMA and @SugaSeanMMA throwing down Saturday at #UFC292!

33 145 710 809.7K



CHILDHOOD IS BEING TAUGHT THAT HITLER WAS THE VILLAIN
ADULTHOOD IS LEARNING THAT HE WAS THE HERO
imgflip.com

@I_Loooove_Pizza, X, 16 August 2023,
https://twitter.com/I_Loooove_Pizza/status/1691995395302314043
@UFC, X, 16 August 2023, https://twitter.com/ufc/status/1691917749914877975

CCDH Center for Countering Digital Hate

© Center for Countering Digital Hate Inc