

Synthetic Media Overview

Synthetic Media is a catch-all term used to describe any media (image, video, text, and/or audio) that is algorithmically created or modified, often using artificial intelligence (AI) techniques. **"Deepfake"** is a type of synthetic media in which an image, video, and/or audio file of a real-life subject has been modified by altering the likeness (usually face) and/or voice of a person and/or replacing it by that of another. A variety of manipulation techniques can be used to generate synthetic media including **Synthesis**, or the creation of realistic media that does not build upon an original piece (other than a prompt) but rather uses algorithms to create a piece of media that is entirely original (e.g., text synthesis, image synthesis). Notably, synthetic media can be enhanced or made more realistic via further manual editing (e.g., photoshopping) of the automated output in order to hide artifacts. These types of media are also known as **Post-Processed Synthetic Media**.



Top: Examples of deepfakes from the Celeb-DF dataset.¹ While the first image is taken from a frame of a real video, the following five images are from deepfake videos using the faces of different subjects.



Left: Example of image synthesis. This portrait image was entirely generated using a deep learning model that was trained on images of people's faces.²

Current GEC Capabilities

GEC Has a Custom AI-Based Algorithm to Detect Synthesized Social Media Profile Pictures of People

GEC has developed an algorithm that can extract thousands of social media profile photos and identify which accounts are using AI-synthesized images of people like the one pictured in the top right.

Profile pictures of people generated via image synthesis are one of the most common types of synthetic media found on social media platforms.

Synthesized profile pictures are used for many reasons:

- by sock-puppet accounts to spread mis/disinformation and propaganda,
- by co-managed accounts to create an appearance of legitimacy, or
- by real users to obfuscate their identities for personal security.



Left: Example of a Twitter account identified in a previous GEC analysis as using an AI-synthesized profile picture. The account published content favorable to the Cuban government, including images associated with Cuban nationalism and propaganda.³

GEC's Next Steps

GEC is investigating techniques to identify deepfake or AI-generated videos on social media. Although GEC has not found significant usage of deepfake or AI-synthesized videos and audios for spreading mis/disinformation and propaganda, the potential for the misuse of deepfakes exists and they continue to grow in popularity on social media. Recent external research alleges the PRC deployed post-processed synthetic media (AI-generated videos) for an online influence campaign. The research claims that this was the first observation of a state-aligned operation promoting videos of AI-generated fictitious people.⁴

GEC is identifying methods for detecting a variety of digital modifications of images posted to social media such as photo editing. GEC currently has access to a third-party tool for identifying some digital modifications on images. However, this tool is unable to perform detections at scale.⁵



Above: AI-generated images of people acting as news anchors and promoting pro-PRC messages in videos by a fictitious media company called Wolf News.⁴

1. Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. (2020). Celeb-df: A large-scale challenging dataset for deepfake forensics. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 3207-3216).
 2. <https://this-person-does-not-exist.com/img/avatar-11090ee4c40291b1aaa29053336a8fa5.jpg>
 3. "Deepfakes Used in Pro-Cuba Networks to Enhance Believability of Disinformation and Propaganda," GEC2022-WHA-721, 27 October 2022
 4. <https://public-assets.graphika.com/reports/graphika-report-deepfake-it-till-you-make-it.pdf>
 5. This tool allows analysis on a single image at a time

Deepfake Meeting: Readout + Proposals

14 Feb Participants: A&R RADS, AFS: Tampa Tech Team & Deepfake SME

27 Feb 2023

Current State of Deepfakes

✓ How our Detection Approach Rates Theoretically

- Pretty well covered for most likely actor behaviors!

✓ Deepfake Videos Not an Imminent Concern

- Less common from adversaries → revisit over next ~6 mos.

✓ More importantly, what are we missing?

- **Photoshopped images (more so)** → item removed, town name changed, etc.
- **Stable Diffusion models (less so)** → text-to-image generative AI model capable of creating photorealistic images given any text input



2023 – Novel instances of AI generated avatars depicting news anchors appeared in Venezuelan, Chinese, and Burkinabe channels. Content reportedly shared via YouTube, Twitter, TikTok, and WhatsApp.

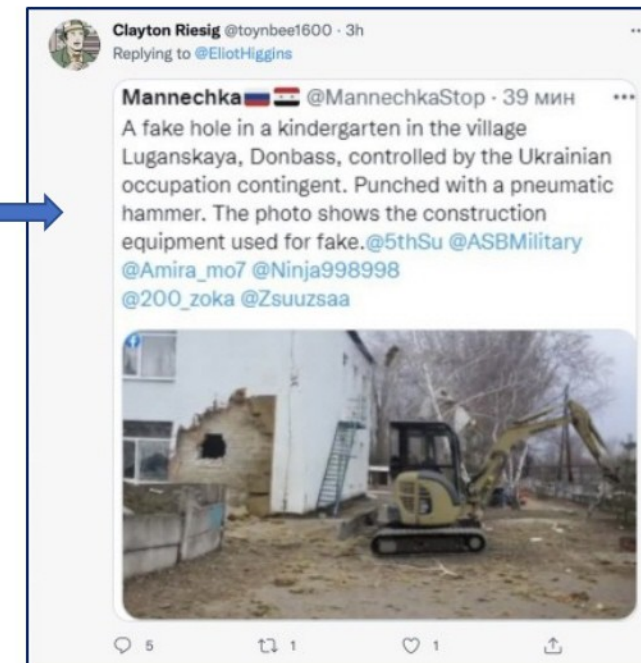
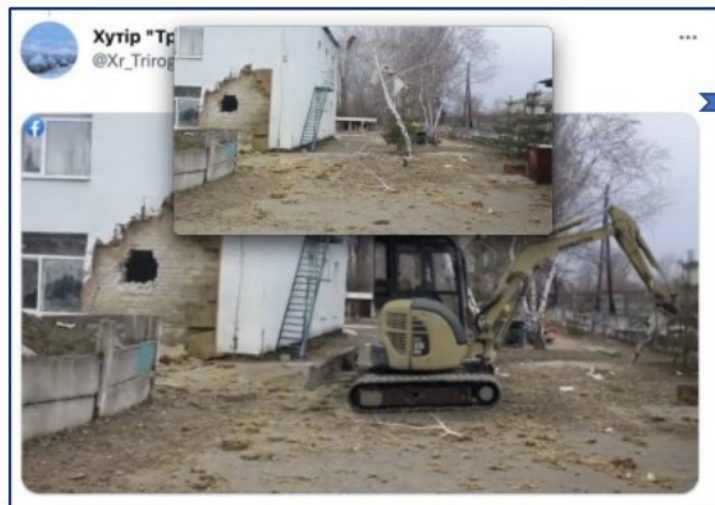
How Prevalent are Photoshopped Images?

F 2025-12348 A-00000828 14 UNCLASSIFIED 11/18/2024

- Non-face image veracity has not been a traditional analysis angle for us
 - Current toolset (image clustering) reveals spread of similar image, but can we determine veracity of posts about images from a reported incident?
- Reflections in Real Life
 - Purported Russian gendered disinformation campaign against German Chancellor candidate Annalena Baerbock in 2021.
 - NIH study that 25% of participants edited >40% of their photos on social media.
 - NYT article found ~4% of scientific papers have altered photos of lab/scientific results.
 - Russian missile strikes and other incidents edited to provide military validity.
 - Kazakhstan editing the visage of new President Tokayev in 2019.

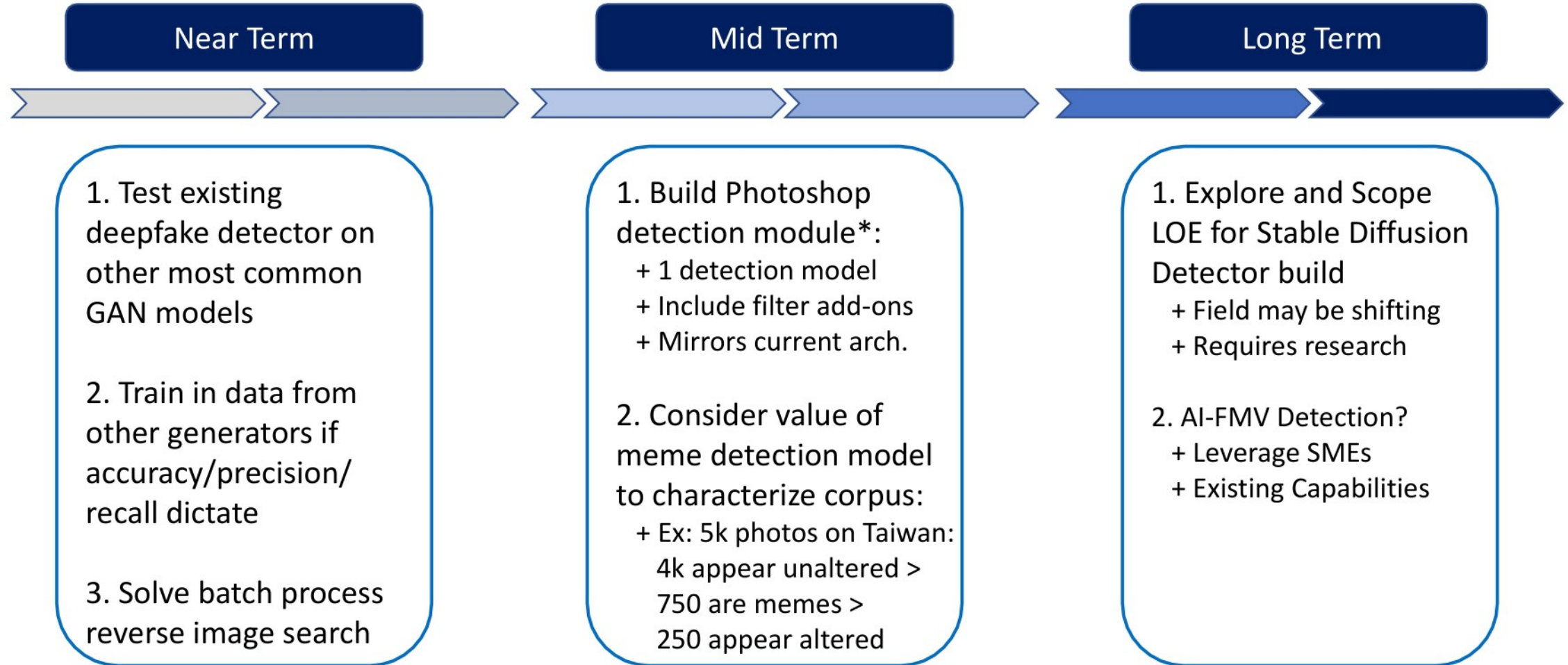


2019 – Kazakh govt photoshopping images of President Tokayev.



2022 – Russian missile strike on kindergarten: photoshopped with construction machinery added to variously justify the kindergarten as a military target, or explain the ‘missile strike’ was faked by the pneumatic attachment

Proposed Next Steps



* Explore internal / broader existing solutions.

Proposal: Photoshopped Images Detection Module

Use Case:

- While A&R retains multiple effective tools to dissect and analyze text-based disinformation content, we do not currently have a structured means for analyzing the veracity of non-face images. Ideally, we will either train a model or develop a suite of 'filters' designed to search for specific image manipulation techniques. This capability would augment and enhance the existing synthetically generated facial detection tool (deepfakes), as well as proposed future development of batch processing reverse image searches.

Background:

- A&R currently conducts image verification by manually running single images through Google Image Search or the Storyzy platform in the hopes of identifying prior usage of the image or other versions where it has not been doctored. While is this somewhat effective for small, curated collections of suspicious images, this methodology does not scale well, relies on human ability to detect image manipulation signatures, and does not produce quantifiable scores conveying confidence in the assessment.

Purpose:

- Create a model or suite of scores (contained in a module) that will receive an image, analyze for common markers of manipulation such as JPEG compression or noise filters, and output a metric(s) indicating likelihood the image has been photoshopped. Architect capability to ensure cohesion with existing and future desired functionalities.

Proposed Approach:

- MVP: ML model that outputs confidence score image is altered/unaltered
 - o Use pretrained image classification model, test for initial performance
 - o Collect training data, sanitize and apply labels
 - o Train and tune parameters as needed

Estimated LOE:

- Number of hours: 40 x 16 weeks = 640
- Time to MVP capability: 3-4 months
- Anticipated Resources: ~3 (1 ML/Architecture engineer, 1 data scientist, 1 analyst)
 - o AWS constraint if/when training model

Proposal: Automated Reverse Image Searches

Use Case:

- A&R increasingly requires the ability to efficiently search hundreds or thousands of images for indications the image has been inauthentically repurposed or fraudulently depicted. This capability would supplement the existing synthetically generated picture detection tool (deepfakes), assist us with observing the spread of an image, and aligns with the proposed future development of a machine learning model trained to detect photoshopped images.

Background:

- A&R currently conducts reverse image searches by manually running single images individually through Google Image Search or a similar platform. While effective for small, curated collections of suspicious images, this methodology does not scale well and combining results from multiple image search platforms is time intensive.

Purpose:

- Automate existing analytic process to solve for efficiency and scale. Architect capability to ensure cohesion with existing and future desired functionalities.

Proposed Approach:

- MVP: scraped images from Twitter can be automatically run through Google or TinEye image search.
 - o Once MVP is working, extend out to other image search services, and then layer in photos without URLs.
- Paid Alternatives:
 - o Storyzy manual URL/image uploads
 - Included in price of subscription
 - o TinEye paid API service:
 - Pay as you go - \$200 for 5,000 image searches (\$.04/search) up to \$10,000 for 1M searches (\$.01/search). Enterprise pricing available too.
 - o Microsoft Bing paid API service:
 - 1,000 per month free (3 transactions per second)
 - \$3 - \$7 per 25,000 transactions (250 txns/second)
 - o PimEyes paid service:
 - \$300 per month for unlimited searches

Estimated LOE:

- Time to MVP capability: 2 months
- Number of hours: 40 x 8 weeks = 80
- Number of Resources: ~2 (1 ML/Architecture engineer, 1 data scientist)

F-2023-12348

A-00000828112

"UNCLASSIFIED"

11/18/2024



(U) GEC Analytics & Research Team Partner Overview

(b)(6)

Director, GEC Analytics & Research
6 June 2023



(U) Intro to Analytics & Research (A&R)



(U) GEC's Analytics & Research (A&R) Division enables Department, Interagency, and International partners to *recognize, understand, and address foreign propaganda and disinformation through methodological, subject matter, and technological expertise*. We provide timely and actionable applied research that is both strategic and tactical to inform and augment efforts in execution of GEC's mission.

(U) Our analytic capabilities include examinations of activity on a variety of social and digital media platforms, synthetic or "deepfake" image and text detection, search engine optimization analysis, and more.

(U) Departmental Core Functions

- (U) Analytics & Audience Insights
- (U) Threat and Regional Integration
- (U) Scientific and AI Policy and Standards
- (U) GEC-IQ Development and Enhancements
- (U) Analytic Coordination

(U) Interagency Core Functions

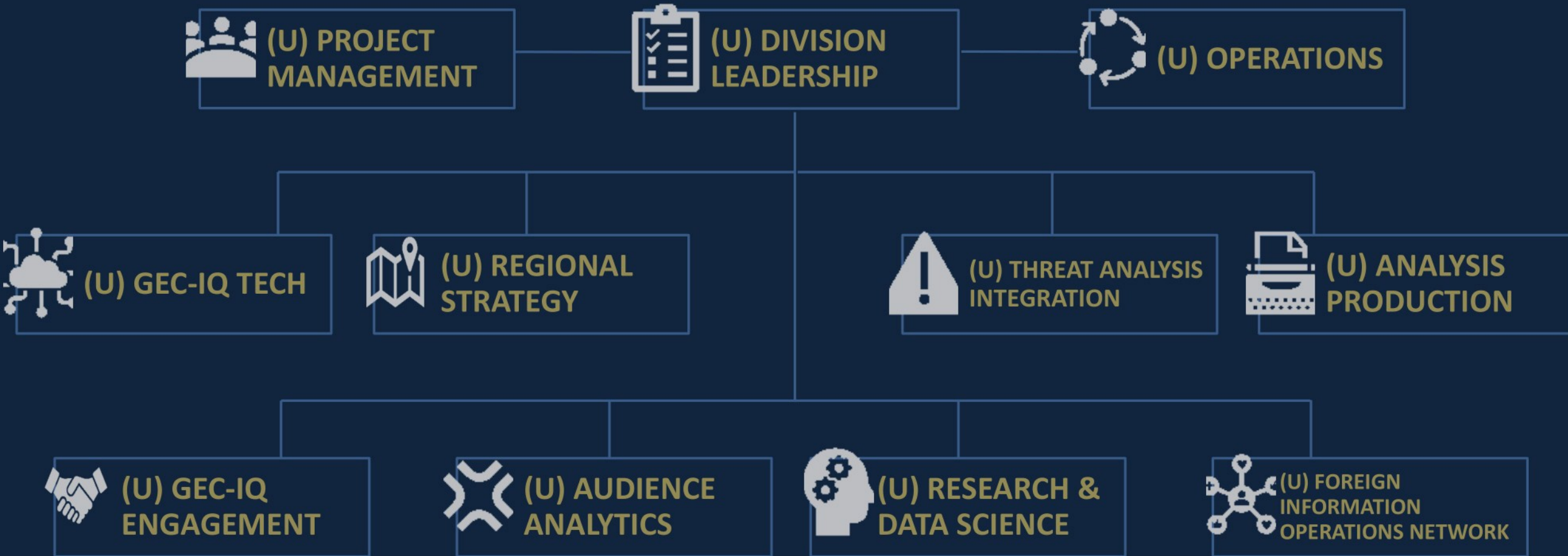
- (U) Analytic Support & Collaboration
- (U) GEC-IQ Deployment & Sustainment
- (U) Technical Leadership
- (U) Information Sharing Best Practices via the Foreign Information Operation Network (FION)

(U) International Core Functions

- (U) GEC-IQ Deployment & Sustainment
- (U) Methodological Upskilling & Collaboration

(U) A&R Division Structure

F-2023-2338 A-000082111 "UNCLASSIFIED" 11/18/2024



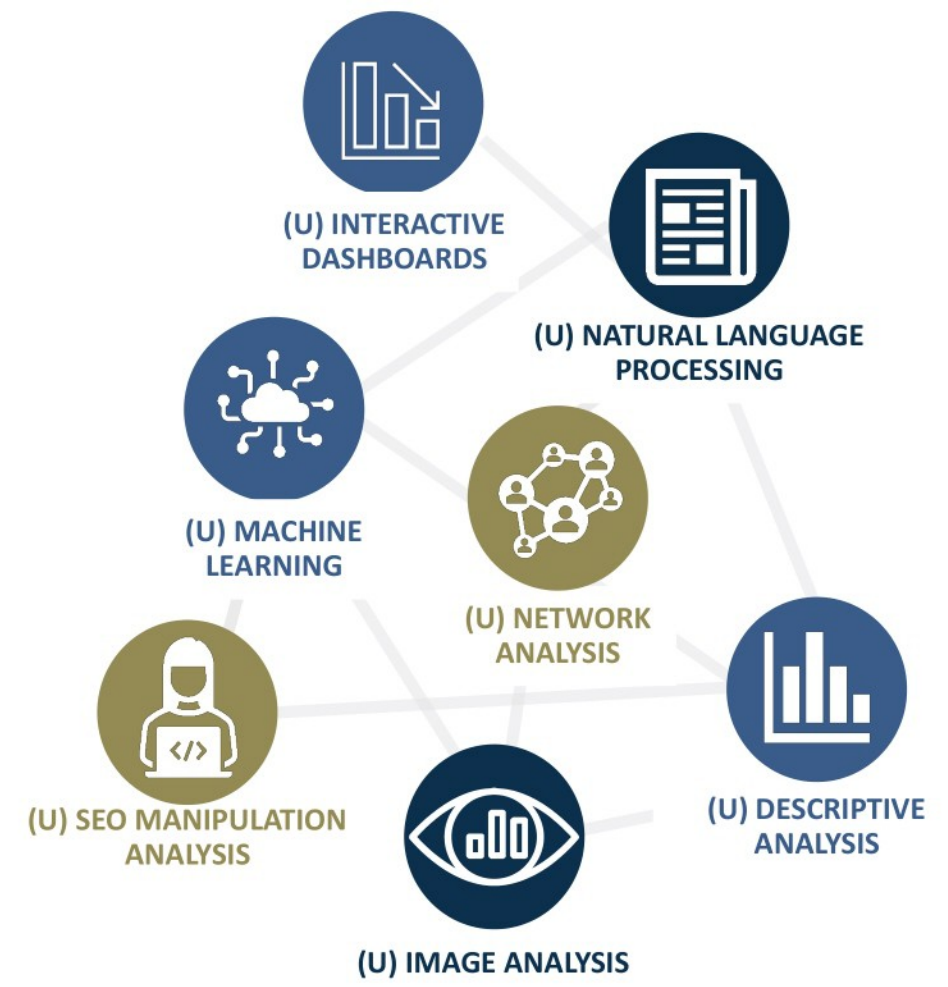
(U) RADS Analytic Capabilities

F:2023-12948 A:000082811 "UNCLASSIFIED" 11/18/2024



(U) Typical Questions

- (U) To what degree was the Twitter conversation shaped by **inorganic activity** (e.g., bots and trolls)?
- (U) What are **public attitudes** in a country toward specific disinformation narratives?
- (U) What impact did a **trending terrorist press release** have on the media environment?
- (U) How is **adversarial propaganda** characterizing an ongoing crisis in traditional media, and what messaging is being promoted through social media channels?
- (U) Are **deepfake images** being used as part of an online messaging campaign?





(U) GEC Analytics & Research Synthetic Media Overview



(U) Synthetic Media Overview



(U) Synthetic Media is a catch-all term used to describe any media (image, video, text, or audio) that is algorithmically created or modified, often using artificial intelligence (AI) techniques.

(U) Deepfake

- (U) An image, video, and/or audio file of a real-life subject that has been modified by altering the likeness (usually face) and/or voice of a person and/or replacing it by that of another.

(U) Synthesis

- (U) Media that does not build upon an original piece; uses algorithms to create media that is entirely original (ex: text synthesis, image synthesis).



Top: Examples of deepfakes from the Celeb-DF dataset.¹ While the first image is taken from a frame of a real video, the following five images are from deepfake videos using the faces of different subjects.



Left: Example of image synthesis. This portrait image was entirely generated using a deep learning model that was trained on images of people's faces.²

(U) RADS in Action: Synthetic Image Detection

(U) RADS has developed an algorithm to detect use of **synthetic imagery within campaigns.**

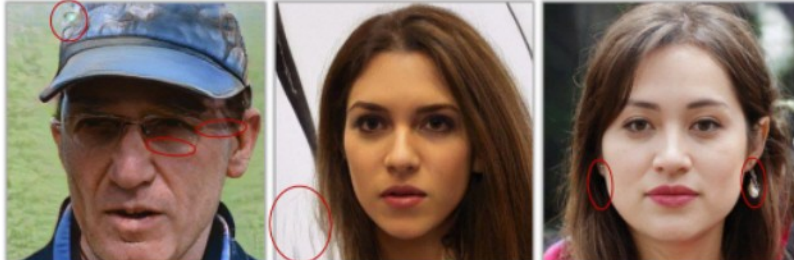
(U) Recent analysis **identified 33 accounts active in two pro-Cuban government amplifier networks** that displayed characteristics of inauthenticity:

- (U) These **highly active accounts** amplified content solely related to pro-Cuban government narratives.
- (U) These accounts also posted images **expressing Cuban nationalism and propaganda** or featuring prominent Latin American leftist leaders or pan-Leftist ideology.
- (U) Further, these accounts' profile and banner pictures **typically included Cuban nationalistic phrases, imagery, emojis.**

UNCLASSIFIED

GEC DEEPPAKE IDENTIFICATION

Deepfakes are synthetic or manufactured, but extremely realistic, images of people. Such images typically have obscure backgrounds, neutral expressions, and aligned eyes. While facial features can be very convincing, when subjected to closer inspection there are many indicators that an image was GAN-generated. Some examples of these indicators seen in this analysis are included below.



Deepfake image for user [@rubioamerica](#), contained an artifact, an incohesive area of color or texture, and struggled to create accurate glasses frames.¹⁸

Deepfake image for account [@rubioamerica2](#) was identifiable by computer-generated hair not connected to a source.¹⁹

GAN-generated images can struggle with symmetry, sometimes creating mis-matched earrings like in this image for [@latinaalexis03](#).²⁰


UNCLASSIFIED

GEC DEEPPAKE PATTERN-OF-LIFE ANALYSIS

The cumulative posting activity of these accounts suggests human operation rather than partial or full automation. The concentration of posts during typical waking hours suggests substantial human involvement, potentially as part of an organized operation funded in whole or in part by one or more actors. Account operators may utilize deepfake profile images as a means of boosting the account's relative credibility (compared to those accounts with no profile image or profile images of flags or prominent figures). This technique may also obscure the identity of the human operator while allowing them to utilize a daily sleep-wake routine.


- Posting volume was steady throughout the week, with accounts posting more during daytime hours than overnight. The highest activity levels were in the evenings (8-10pm).
- The overall activity pattern resembles real individual users rather than mostly automated accounts or accounts that are operated as a result of a full-time, 24/7 state-sponsored disinformation or propaganda campaign, resulting in greater activity during typical "business" (or waking) hours.
- Deepfake profile images could be used to provide an additional element of credibility while simultaneously allowing the operator to remain anonymous, as human-appearing profile images are typically deemed more trustworthy by other users online.

DEEPPAKE ACCOUNTS' DAILY POSTING PATTERN



Day of Week	Tweet Volume
Monday	~18,000
Tuesday	~18,000
Wednesday	~18,000
Thursday	~18,000
Friday	~18,000
Saturday	~18,000
Sunday	~18,000

DEEPPAKE ACCOUNTS' HOURLY POSTING PATTERN



Hour of Day (EST)	# of Tweets
12:00 AM	~2,000
1:00 AM	~1,000
2:00 AM	~1,000
3:00 AM	~1,000
4:00 AM	~1,000
5:00 AM	~1,000
6:00 AM	~1,000
7:00 AM	~2,000
8:00 AM	~4,000
9:00 AM	~8,000
10:00 AM	~10,000
11:00 AM	~8,000
12:00 PM	~6,000
1:00 PM	~6,000
2:00 PM	~6,000
3:00 PM	~6,000
4:00 PM	~6,000
5:00 PM	~6,000
6:00 PM	~6,000
7:00 PM	~6,000
8:00 PM	~12,000
9:00 PM	~12,000
10:00 PM	~10,000
11:00 PM	~6,000

UNCLASSIFIED

(U) Actor Preference : Manipulated Images



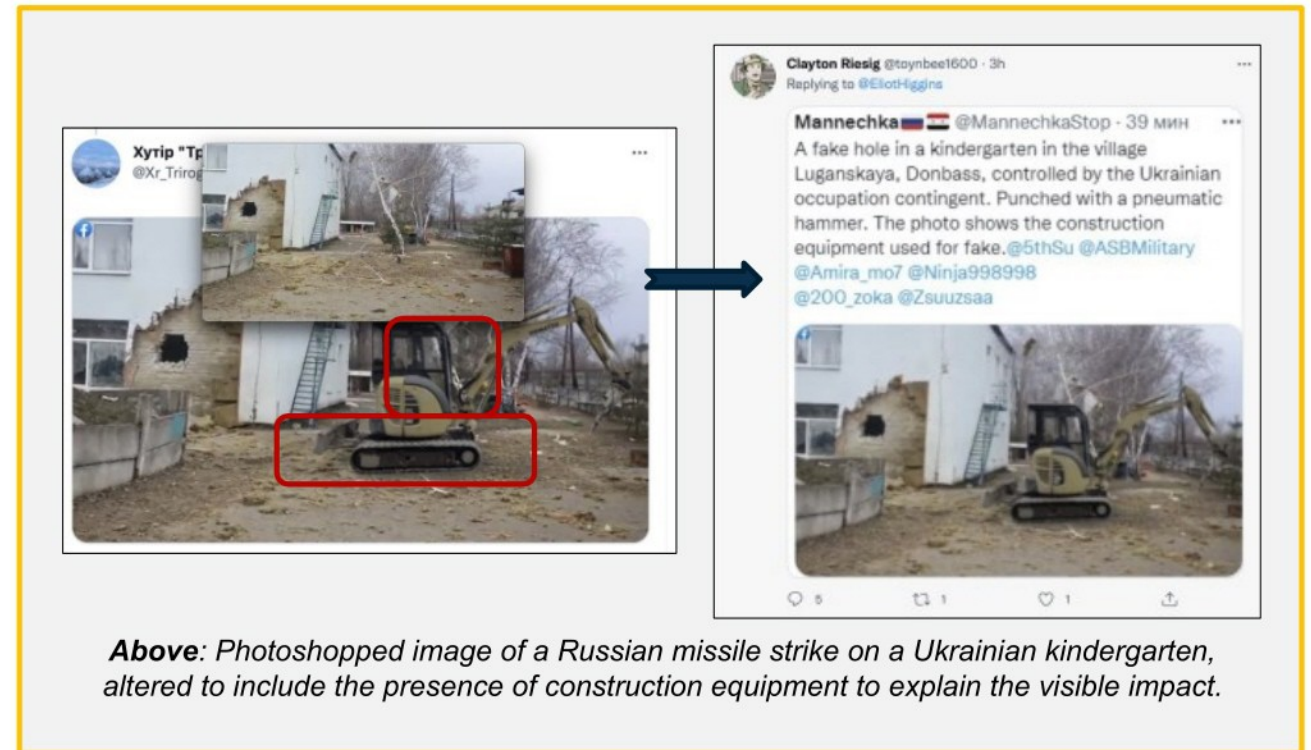
(U) RADS has observed the use of longstanding image manipulation techniques, which are usually low effort attempts to shape event perception.

(U) Photoshopping Images

- (U) Open source tools exist.
- (U) Manipulated images are usually detectable with the human eye, and/or suite of algorithms.
- (U) Most common on social media where rapidly spreading imagery is difficult to refute later.

(U) Use cases

- (U) Disproving misappropriated or fake imagery.
- (U) Identifying altered images, to provide awareness of actor's sensitivities.



(U) Over the Horizon Video and Audio



(U) Synthetic media is rapidly evolving to blur the lines between reality and fully synthetic content.

(U) Current Environment

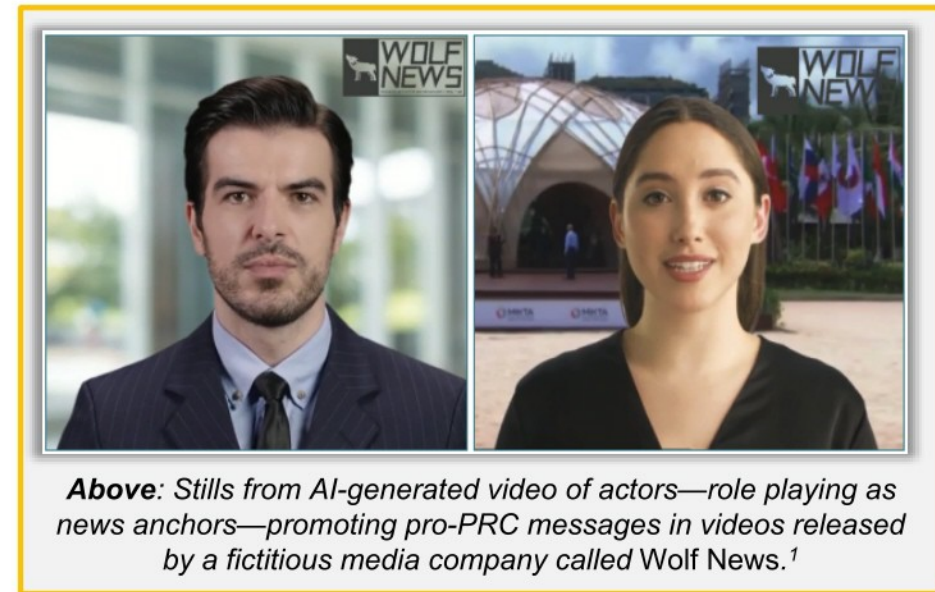
- (U) Usage of deepfake or AI-synthesized videos and audio deemed uncommon for now.

(U) Evolutions

- (U) Recent external research alleges the PRC deployed post-processed synthetic media (AI-generated videos) for an online influence campaign.¹
- (U) New deepfake technologies will require less domain expertise, less technical ability, and less time and can scale rapidly while maintaining a high level of believability.

(U) Over the Horizon

- (U) Synthetic audio is evolving rapidly, and could be deployed effectively in countries with high news radio consumption.
- (U) Detection models are lagging, while citing incorrectly flagged material may diminish trust in detection.



(U) Over the Horizon Imagery



(U) AI art generation models are now able to produce hyper-realistic imagery that does not carry markers of manipulation, only its synthetic creation.

(U) Current Environment

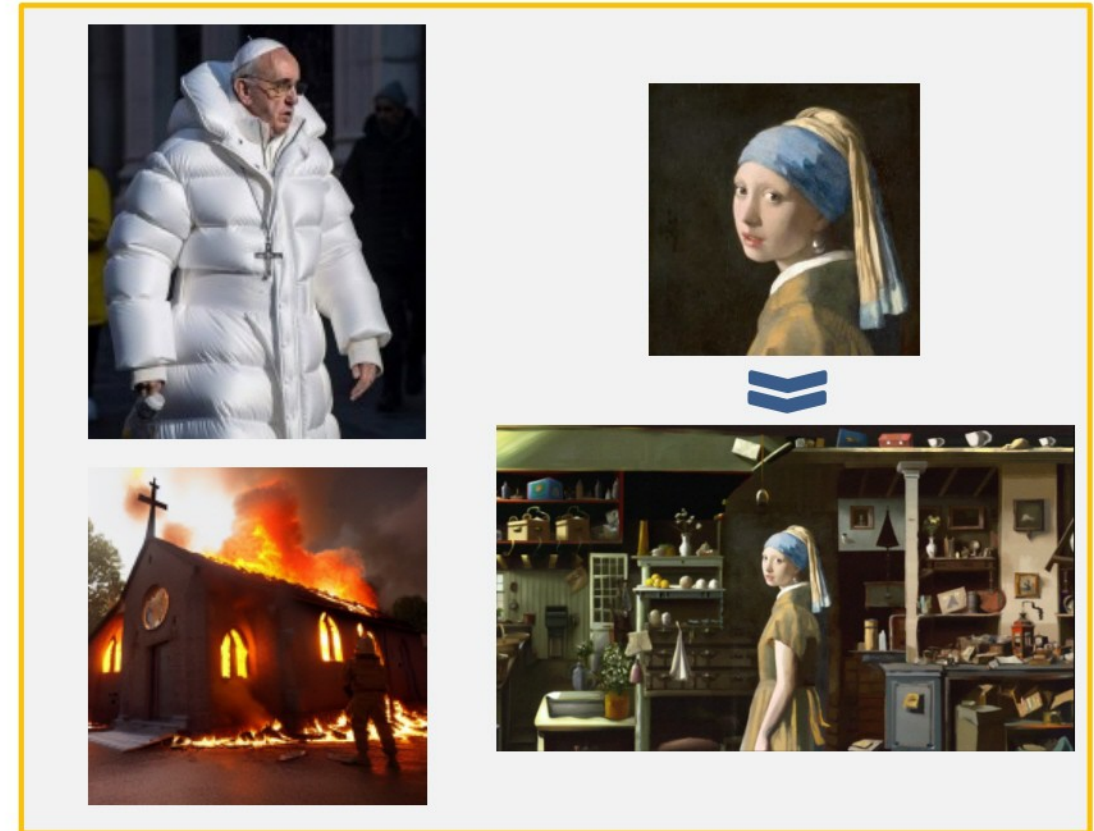
- (U) Usage is primarily in private sector for now.

(U) Evolutions

- (U) Models can create original, realistic images from a text prompt, combining concepts, attributes, and subjects.
- (U) 'Inpainting' and 'outpainting' techniques can augment a real image to add or remove content not originally there.

(U) Over the Horizon

- (U) Synthetic images could be deployed in information campaigns with specific, targeted goals.
- (U) The main guardrails are developers' self-imposed input constraints; will that change?



Clockwise, from Top Left: 1. photo-realistic AI generated image; 2. original image, 'outpainted' by AI model to include a broader synthetic scene; 3. AI generated, photo-realistic image of a burning church.

Pro-Ortega-Murillo Regime Amplification Network Exhibits Heightened Focus on Internal Audiences Instead of Regional and Extra-Regional Adversarial Narratives

COORDINATED COMMUNITY DETECTION & SOCIAL MEDIA ANALYSIS

EXECUTIVE SUMMARY

Analysis of conversations in advance of Nicaragua's 6 November 2022 municipal elections suggests that neither deepfakes nor coordinated inauthentic behavior were key components of the pro-Ortega-Murillo amplification network's tactics. Instead, accounts tended to repeat similar, specific content that garnered relatively low engagement across the platform. These tactics differ greatly from those utilized by other core regional amplification networks in Latin America when amplifying internally focused messaging—specifically those operating at the behest of the Maduro regime in Venezuela and the Cuba government. Given the behaviors exhibited by this network in previous GEC reporting, the pro-Ortega-Murillo network seems to exhibit a greater degree of tactical variance, utilizing different procedures when focusing internally and externally, than other regional amplification networks.

OBJECTIVE

This analytical effort seeks to advance the U.S. government's understanding of the tactics, techniques, and procedures (TTPs) utilized on Twitter by the Ortega-Murillo regime in Nicaragua, and/or their witting or unwitting amplifiers. We examined conversations posted in in advance of the 6 November 2022 municipal elections to understand the way in which the regime seeks to alter and shape perceptions within Nicaragua, and the extent to which such efforts resonate in the information space.

KEY FINDINGS

- The Ortega-Murillo regime differs tactically from Maduro regime and Cuban amplification networks in two key aspects: 1) the disinformation and propaganda created, disseminated, and amplified by the network focuses on shifting perceptions within the Nicaraguan domestic population; and 2) in contrast to other regional networks, there is a less overt state involvement in the production, dissemination, and amplification of disinformation and propaganda content.^{A, B}
- At least four of 22 broadcaster accounts¹ situated in the identified Ortega-Murillo regime amplification network regularly displayed behaviors indicative of inorganic amplification that violate the statistical assumptions about "Normative User Account Behavior."²
- GEC identified 26 accounts within the network that utilized deepfake profile images, half of which were previously identified in our assessment of the network benefiting the Cuban government, entering the network mainly as a result of being tagged or retweeted by accounts in the Ortega-Murillo network. Despite their inclusion in this network, the 13 Cuba-focused accounts did not appear to prioritize disseminating or amplifying Ortega-Murillo regime themes, suggesting these accounts retained their focus on Latin American pan-Leftist narratives.^C

¹ Broadcasters are a group of highly active, interconnected accounts that are involved in the targeted dissemination and amplification of specified content. In this case, the amplification and dissemination of pro-Ortega-Murillo regime and pro-FSLN content.

² Our assumptions about normative user account behavior are based on constraints imposed by the human condition. As a result, account behavior patterns that indicate a lack of sleep, publishing tweets within seconds of one another in such a manner that exceeds human capacity, or sharing statistically large volumes of tweets on a daily basis such that it is incongruous with the behavior of other accounts in the network are all violative of normative user account behavior.

BACKGROUND

This report serves to advance understanding of the TTPs utilized by the primary regional actors of concern in the disinformation and propaganda space. Prior GEC reporting studied the networks associated with the Cuban government and the Maduro regime in Venezuela, but coverage of efforts benefitting the Ortega-Murillo regime has largely been ancillary. Accordingly, the 6 November 2022 Nicaragua municipal elections provided an opportunity to examine efforts to shape, influence, and manipulate the domestic information space, particularly given that the network operating to the benefit of the Ortega-Murillo regime tends to primarily focus on the targeted dissemination and amplification of disinformation and propaganda directed at audiences internal to Nicaragua.

The networks benefitting the Cuban government or the Maduro regime in Venezuela appear to focus both internally and externally, though our analyses have found that they devote greater effort to attempting to shape external perceptions. Predictably, content produced, disseminated, and amplified by these networks tends to focus on narratives that we categorize as being general Latin American pan-Leftist content, including: 1) anti-sanctions/anti-embargo narratives; 2) general anti-U.S./anti-Western narratives (which may often echo Russian or PRC talking points); and 3) region-specific, anti-imperialist, and anti-neocolonialist narratives (which largely target and criticize the United States). However, while previous GEC analyses have confirmed that accounts generally focused on the Nicaraguan information space do participate in these broader, pan-Leftist conversations, our findings suggest that most of the content produced in the network benefitting the Ortega-Murillo regime focuses on pro-Ortega-Murillo regime propaganda content, providing positive portrayals of the regime's governance and life in Nicaragua.

GEC is unable to assess, based on open-source data, that the accounts identified in this analysis and those driving the campaigns benefitting the Ortega-Murillo regime are controlled by the regime (which would be analogous to our findings regarding the Cuban and Maduro regime-aligned networks). Analysis of the behavior of the accounts in the network, including their focus on the production, dissemination, and amplification of content favorable to the Ortega-Murillo regime and/or (to a lesser extent) ideologically aligned governments in the region leads to the probabilistic conclusion that the Ortega-Murillo regime may have some role in this network. However, GEC has not uncovered clear evidence (to include affirmative, self-described attribution as a government account, official accounts associated with government ministries, or official or personal accounts utilized for official purposes associated with high-ranking officials in the government or regime) of control or direction by regime officials as we have in the case of Cuba or the Maduro regime. Accordingly, we continue to describe this network as operating to the benefit of, not at the behest or direction of, the Ortega-Murillo regime.

REPORT

Analysis of the content and metadata of 290,000 tweets published in advance of the 6 November 2022 municipal elections in Nicaragua uncovered 26 accounts utilizing deepfake³ profile images, with half of those accounts having been identified in our previous analysis of a Twitter-based information environment manipulation network linked to the Cuban government. Additionally, two broadcaster accounts identified in our analysis, @FelipeC55155489 and @Nbalmaceda2, exhibited multiple hallmarks of likely inauthenticity, including the use of generic profile pictures⁴ as well as multiple behaviors that violate our standard assessment of normative user account behavior. Although we observed the prevalence of propaganda narratives in the Nicaraguan information space that we assess to be of benefit to the Ortega-Murillo regime, we did not observe large-scale coordinated manipulation of the information environment in advance of the municipal election, nor did we observe the broad use of deepfake profile images to advance pro-regime narratives resonating in the information environment.

LIKELY INAUTHENTIC ACCOUNTS UTILIZE ESTABLISHED PLAYBOOK TO PROMOTE PRO-REGIME CONTENT IN ADVANCE OF ELECTION

We collected a dataset of 290,000 tweets published between 1 July and 15 November 2022, encompassing the lead-up to the municipal elections in Nicaragua, the 6 November 2022 elections themselves, and the immediate aftermath thereof. Evaluation of this dataset for evidence of coordinated amplification of content or narratives uncovered several accounts engaged in behavior that we classify as artificial or indicative of the presence of artificial components in the account's administration. For example, although the behaviors we observed do not meet the threshold required to determine that the network as a whole, or in substantial part, engaged in coordinated activity, several accounts published tweets up to 24 hours a day, a social media behavior pattern that is incongruous with the profile of typical human users. Nevertheless, we identified several prominent TTPs that we have identified as regional hallmarks of information domain influence, including: 1) the repetition of specific, target content across an array of accounts despite low engagement with the content itself; 2) the publication and amplification of content favorable to the regime at rates inconsistent with free and independent media; and 3) the use of cross-network alliance-building⁵ tactics as well as mention trains⁶ to enhance the visibility of, and increase the probability of amplification of, narratives and content favorable to the Ortega-Murillo regime, including propaganda content presenting an idealized view of Nicaraguan society.

³ A deepfake uses image, video, text, and/or audio that is created or modified algorithmically, often using artificial intelligence (AI) techniques, in which a real-life subject has been modified by altering the likeness (usually face) and/or voice of a person and/or replacing it with that of another.

⁴ We refer to profile photographs as "generic" when they are, or appear to be, stock photography, animals, photographs of famous or public individuals (provided it is not the account owner) associated with the government or regime being supported, and/or blank or single color backgrounds.

⁵ Alliance-building is the tactic by which accounts tag identified accounts most commonly associated with other amplification networks (usually those accounts that are generally within the amplification networks associated with the Maduro regime or the Cuban government) in what we assess to be an effort to encourage ideological allies to disseminate and amplify the targeted message.

⁶ A mention train is a tweet or series of tweets that contain nothing but a list of user names, some emojis, and usually a meme or a gif which may or may not have a direct relation to the message being amplified.

Within our set of broadcaster and seed accounts, a few stood out for their unique account behavior patterns; specifically:

- broadcaster account @yadiratellez4 regularly published between four and six tweets in a single minute, and averaged between 50 and 60 tweets per day;
- broadcaster account @Nbalmaceda2 at times published tweets as frequently as one every seven and a half seconds, while averaging between 40 and 60 tweets per day; and,
- broadcaster accounts @Atego16 and @FelipeC55155489 published tweets 24 hours a day, and accumulated average daily tweet volumes similar to the accounts @yadiratellez4 and @Nbalmaceda2.

Notably, while accounts @yadiratellez4 and @Nbalmaceda2 accumulated high average daily tweet volumes, both increased the hours during which they published content to 24 hours per day in the immediate lead-up to the election. We assess this change in behavior to be indicative of an attempt to increase the prevalence of, and enhance the dissemination and amplification of, pro-Ortgea-Murillo regime content in the pre-election period to provide a veneer of broad popular support for the regime.

Account	Type	Identified TTPs
@yadiratellez4	Broadcaster	This account exhibited a high daily volume of published tweets atypical for the prototypical human user; this account exhibited a rate of publication that is inconsistent with human user capabilities; this account published tweets up to 24 hours a day prior to the election, a usage period that is not consistent with a normal human user
@Nbalmaceda2	Broadcaster	This account exhibited a high daily volume of published tweets atypical for the prototypical human user; this account exhibited a rate of publication that is inconsistent with human user capabilities; this account published tweets up to 24 hours a day prior to the election, a usage period that is not consistent with a normal human user
@Atego16	Broadcaster	This account exhibited a high daily volume of published tweets atypical for the prototypical human user; this account exhibited a rate of publication that is inconsistent with human user capabilities; this account published tweets up to 24 hours a day prior to the election, a usage period that is not consistent with a normal human user
@FelipeC55155489	Broadcaster	This account exhibited a high daily volume of published tweets atypical for the prototypical human user; this account exhibited a rate of publication that is inconsistent with human user capabilities; this account published tweets up to 24 hours a day prior to the election, a usage period that is not consistent with a normal human user
@NicaSoberana	Seed	This account published tweets between 22 and 24 hours a day, a usage period that is not consistent with a normal human user.

F-2023-12348

A-00000828110

"UNCLASSIFIED"

11/18/2024

DEEPPAKES NOT INTEGRAL TO PROPAGANDA BENEFITING THE ORTEGA-MURILLO REGIME

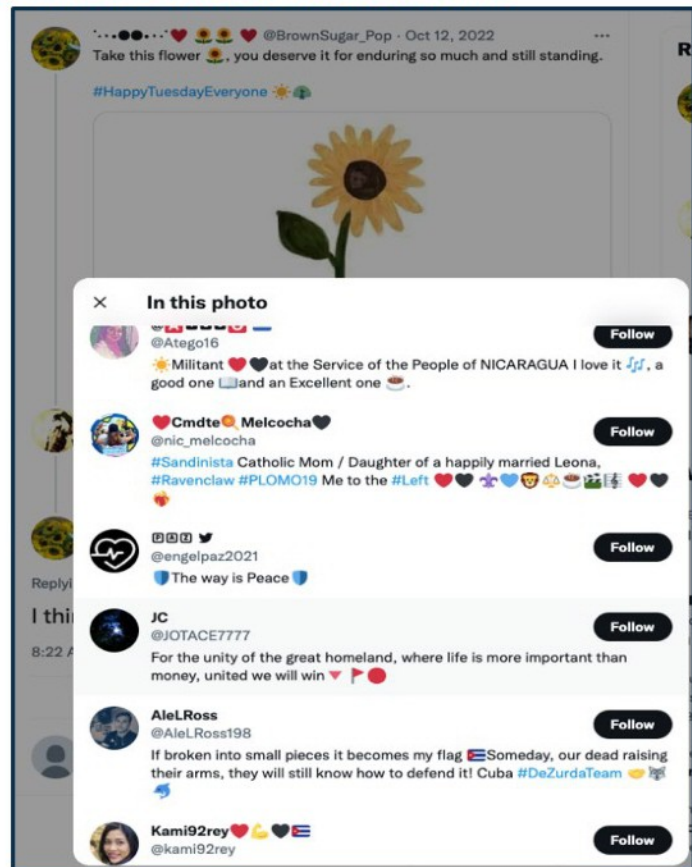
Of the 26 accounts in our dataset that we identified as using deepfakes, only four appear to have locations in Nicaragua (based on self-reported locations and activity histories). These four accounts appear to have had a negligible impact on the Nicaragua information environment, combining for a total of 41 tweets during the reporting period. Half (13) of the 26 accounts that utilized deepfakes were accounts that we previously identified in our assessment of a network of accounts disseminating and amplifying content favorable to the Cuban government. The majority of these 13 accounts appear to have been included in our dataset as a function of alliance-building accounts mentioning the accounts as part of mention-train tweets that also include broadcaster accounts in the pro-Ortega-Murillo amplification network.

Prior GEC analyses have evidenced Nicaragua-focused accounts amplifying content initially produced and disseminated by the Cuba amplification network, particularly when the content promotes pan-Leftist narratives.

Accordingly, we assess that this tagging effort may have been an effort to draw upon the resources of the Cuba amplification network to strengthen dissemination efforts. Those efforts do not appear to have been successful. Once tagged into the network, Cuban deepfake accounts published a total of only 140 retweets and 22 replies. Moreover, three of the identified 13 Cuba-focused deepfake accounts that were tagged into the pro-Ortega-Murillo network published an average of 9,300 tweets each during the reporting period, although on average only four tweets engaged, and were relevant to, the Nicaraguan information space.

DISCUSSION

The pro-Ortega-Murillo digital amplification network on Twitter in Nicaragua appears to exhibit distinct behavior patterns when compared to other prominent disinformation and propaganda amplification networks in the Latin America. Unlike networks that we have previously assessed (i.e., the Venezuela-based Maduro regime amplification network and the Cuban government amplification network), the pro-Ortega-Murillo amplification network does not seem to rely on coordinated inauthentic amplification techniques to disseminate messaging, nor do deepfakes appear to play a substantial role in the network, at least insofar as it relates to propaganda efforts internal to Nicaragua. One potential explanation for this is that we have yet to see, in open-source reporting, evidence of compensation for amplification as we have with the networks associated with the Maduro regime and the Cuban government. Accordingly, there does not appear to be the same widespread financial inducement to amplify pro-regime narratives in Nicaragua.



An example of a tweet passively tagging Cuban deepfake @kami92rey as present in this photo alongside Nicaraguan seed @nic_melcocha and broadcasters @Atego16 and @superfrog89. Note the non-political nature of the photo.

GEC may consider future research assessing broader coordination efforts between regional amplification networks. To date, our research has examined coordination and cooperation around discrete campaigns or events. While useful, such assessments provide limited views into the broader state of coordination across the associated networks. Additionally, deepfakes (both images and video) are becoming an increasingly prominent vector for the injection of disinformation and propaganda. As a result, GEC may consider future research assessing the intentions underpinning the differentiated usage of deepfakes across the core regional amplification networks.

METHODOLOGY

Previously, GEC conducted a combined account and hashtag social network analysis that revealed a set of “seeds”—accounts with an established history of publishing propaganda favorable to the Ortega-Murillo regime—and “broadcasters”—accounts for which we have been unable to confirm government affiliation in open-source information that nevertheless are highly active in the dissemination and amplification of pro-Ortega-Murillo regime and pro-Sandinista content. This analysis studies the network of 20,194 accounts that interact⁷ with the 22 seed and broadcaster accounts. To identify accounts using deepfake images, we then analyzed the profile pictures systematically leveraging GEC’s custom deepfake detection model which has been trained to discover altered images.⁸

To conduct Coordinated Community Detection (CCD), we cataloged and analyzed the posting patterns of accounts interacting with Nicaraguan seed and broadcaster accounts in the run-up to the 6 November 2022 municipal elections to identify suspicious, inauthentic patterns embedded in both tweet content and metadata. While this analysis included 290,000 tweets and 77,000 images extracted from the tweets, the image analysis portion did not uncover coordinated posting, as results only highlighted the most retweeted images (these findings were addressed in our previous analysis).^D

LIMITATIONS

This analysis only examines open-source, public tweets published between 1 July and 15 November 2022 collected via our analytic platforms. As with all digital collection platforms, there are inherent gaps, and our data may not be fully representative of the digital conversation.

⁷ For the purposes of this paper, we define interaction as a reply, retweet, or mention/tag.

⁸ “Deepfakes” refer to images generated through a family of machine learning models, with generative adversarial networks (GANs) being the most widely used today. Deepfakes are computationally heavy to produce, and GAN-created images are the lightest lift.

[We would appreciate your feedback by completing a short survey here.](#)

REFERENCES

^A For previous GEC reporting detailing the message amplification tactics used by the Cuban government, please see *"Cuban Government's Propaganda and Disinformation Efforts on Twitter Tied to Pan-Latin America Leftist Support and Cuban Medical Brigades, with Tactics Reminiscent of the Maduro Regime,"* GEC2022-WHA-720, 22 September 2022; *"Deepfakes Used in Pro-Cuba Networks to Enhance Believability of Disinformation and Propaganda,"* GEC2022-WHA-721, 27 October 2022; and *"Cuban Propaganda and Disinformation Networks Evidence Sub-Communities Amplifying Targeted Narratives,"* GEC2022-WHA-722, 2 November 2022.

^B For previous GEC reporting regarding the Maduro network, please see *"Maduro Regime Uses Daily Messaging Campaign to Artificially Shape the Information Environment in Venezuela,"* GEC2020-WHA-820, 22 December 2020; *"Maduro Regime Disinformation and Propaganda Amplification Efforts on Social Media Centralized Around a Limited Set of Accounts,"* GEC2020-WHA-821, 22 December 2022; *"Effective Amplification of the Maduro Regime's Messaging Relies Upon a Handful of Well-Established Accounts,"* GEC2020-WHA-822, 22 December 2020; and *"The Tactics, Techniques, and Procedures of Maduro Regime Social Media Message Amplification,"* GEC2021-WHA-823, 4 May 2021.

^C Ibid.

^D For Previous GEC analysis regarding the Nicaraguan amplification network, see *"Influential 'Broadcaster' Twitter Accounts Employ Novel Approach to Bolster Ortega-Murillo Regime and FSLN, Using Unconstrained Narratives Ahead of Election,"* GEC2022-WHA-322, 15 December 2022.



Deepfakes Used in Pro-Cuba Networks to Enhance Believability of Disinformation and Propaganda

Deepfake Images Likely Here to Stay

Analytics & Research

Global Engagement Center

U.S. Department of State

27 October 2022

GEC2022-WHA-721

EXECUTIVE SUMMARY

Analysis of more than 33,000 accounts active in disinformation and propaganda networks that create, disseminate, and amplify content favorable to the Cuban government identified at least 33 unique accounts (0.1% of total accounts assessed) using a deepfake (computer generated) profile image.¹ As a function of total accounts assessed, deepfake profile images are roughly twice as common in this data set as they are in another recent analysis assessing the prevalence of deepfakes in the Russia-Ukraine information space.²

The profiles identified as using deepfake images as account profile pictures also tended to publish images associated with Cuban nationalism or Cuban propaganda (e.g., images of Fidel Castro with nationalistic quotes superimposed), or other images associated with prominent Latin American leftist leaders or Latin American pan-Leftist ideology. Within the two distinct networks associated with Cuban disinformation and propaganda that we have identified and analyzed (the De Zurda Team and the @cubacooperaven networks), all 33 accounts interacted with accounts associated with the De Zurda Team, and 18 of 33 accounts interacted with the @cubacooperaven account.

Notably, the accounts utilizing deepfake images did not show significant signs of automation; instead, they exhibited behavior patterns more akin to those of human-run accounts. For example, we observed that accounts with deepfake profile images tended to be more likely to reply to tweets than other accounts in the networks (as opposed to publishing, retweeting, or quote tweeting content). We believe that this may be indicative of an evolution in operational techniques designed to decrease the likelihood of account suspension.



*De Zurda Team accounts include both @dezurdateam_ (active 1 June 2022 - current) and @dezurdateam (suspended).

DEEPAKE ACCOUNT ANALYSIS

Of the 33 accounts using deepfake profile images, 26 also published content of some form expressing solidarity with the Cuban government, nationalism, or alignment with Cuban government propaganda or disinformation narratives. In some cases, accounts reported being in Cuba, though such claims were not independently verified. Accounts lacking this branding tended to be only minimally active in the identified networks, generally focusing on issues and regions external to Cuba.

- We analyzed account-level information for all 33 accounts identified as using a deepfake profile image. For three accounts that were suspended by Twitter, we only analyzed metadata.
- Within the 30 active accounts—at the time of this analysis—biography and profile contents often identified users as [Cuban patriots](#), and they referenced the Cuban [revolution](#), [socialism](#), and Cuban [political figures](#).³ Four of the 30 accounts lacked any such reference.
- These four accounts were not engaged in relevant Cuban conversations, instead focusing on other geographic locations. For example, one account appeared focused on [Russia](#) and even tweeted in Russian; another account focused on [Spain](#), Catalonia in general and Barcelona specifically; the third account focused on [Brazil](#); while the fourth account focused on [Argentina](#).^{4,5}
- On average, the 26 Cuba-focused deepfake accounts were 2.3 years old and posted 23.2 tweets per day throughout their lifetime (16.7 tweets per day during the analysis period). Additional analysis on account activity is included in the following slides.



Examples above of two account profiles* that used deepfake profile images. Account descriptions and profile imagery often identify as Cuban and reference political and national figures. One bio contained a reference to account suspensions.

*Account profiles were machine translated to English from Spanish.

DEEPPFAKE CONTENT ANALYSIS

In general, accounts that utilized deepfake profile images—notwithstanding exceptions noted previously—promoted content consistent with Cuban government messaging that was favorable to the Cuban government and expressed nationalistic sentiment. Hashtags praised Cuban medical diplomacy efforts, highlighted historical events, and promoted political causes. Centering content around hashtags is a tactic we have previously attributed to the Maduro regime in Venezuela and have noted in our previous reporting related to Cuba.

- We analyzed 156,775 tweets that were published between 1 January and 23 September by the 30 non-suspended accounts that utilized a deepfake profile image.
- Content included in these tweets was heavily focused on leftist Cuban politics. Hashtags related to the De Zurda Team (#dezurdateam), as well as those associated with Cuban medical diplomacy propaganda (#cubaporlavida, #cubaporlapaz) were among the most frequently used.
- Content also included posts about major events in Cuba's revolutionary history, praise for first responders (such as those involved in Cuba's medical brigades), and celebrations of International Workers' Day. Over 1,600 tweets (1.1%) were tagged with #eliminaelbloqueo, offering critiques of the U.S. embargo both in text and in images.

Top 10 Hashtags	Translation	# of Tweets	Percent of Tweets
#cuba	Cuba	36,914	23.5%
#cubaporlapaz	Cuba for peace	12,004	7.7%
#vamoscontodo	Let's give it our all	10,244	6.5%
#cubavive	Cuba lives	8,454	5.4%
#cubaviveytrabaja	Cuba survives and thrives	4,962	3.2%
#dezurdateam	De Zurda Team	4,787	3.1%
#cubaporlavida	Cuba for life	4,773	3.0%
#fidel	Fidel	4,446	2.8%
#fidelporsiempre	Fidel forever	4,366	2.8%
#cubaviveensuhistoria	Cuba lives in its history	4,040	2.6%



Examples of Cuban political and nationalistic content* published by accounts using a deepfake profile image.⁶

*Content was machine translated to English from Spanish.



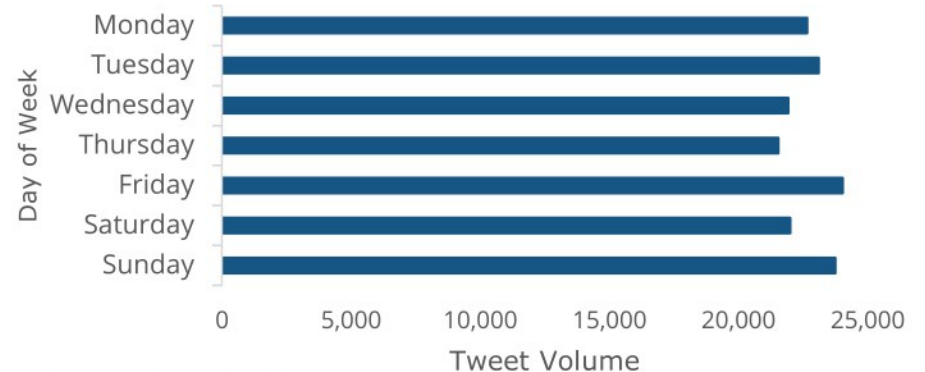


DEEFAKE PATTERN-OF-LIFE ANALYSIS

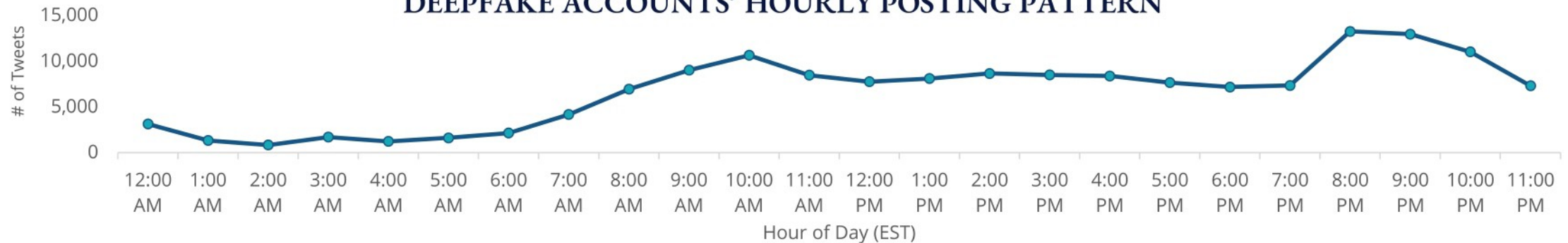
The cumulative posting activity of these accounts suggests human operation rather than partial or full automation. The concentration of posts during typical waking hours suggests substantial human involvement, potentially as part of an organized operation funded in whole or in part by one or more actors. Account operators may utilize deepfake profile images as a means of boosting the account's relative credibility (compared to those accounts with no profile image or profile images of flags or prominent figures). This technique may also obscure the identity of the human operator while allowing them to utilize a daily sleep-wake routine.

- Posting volume was steady throughout the week, with accounts posting more during daytime hours than overnight. The highest activity levels were in the evenings (8-10pm).
- The overall activity pattern resembles real individual users rather than mostly automated accounts or accounts that are operated as a result of a full-time, 24/7 state-sponsored disinformation or propaganda campaign, resulting in greater activity during typical "business" (or waking) hours.
- Deepfake profile images could be used to provide an additional element of credibility while simultaneously allowing the operator to remain anonymous, as human-appearing profile images are typically deemed more trustworthy by other users online.

DEEFAKE ACCOUNTS' DAILY POSTING PATTERN



DEEFAKE ACCOUNTS' HOURLY POSTING PATTERN





DEEFAKE CUBAN NETWORK ACTIVITY COMPARISON

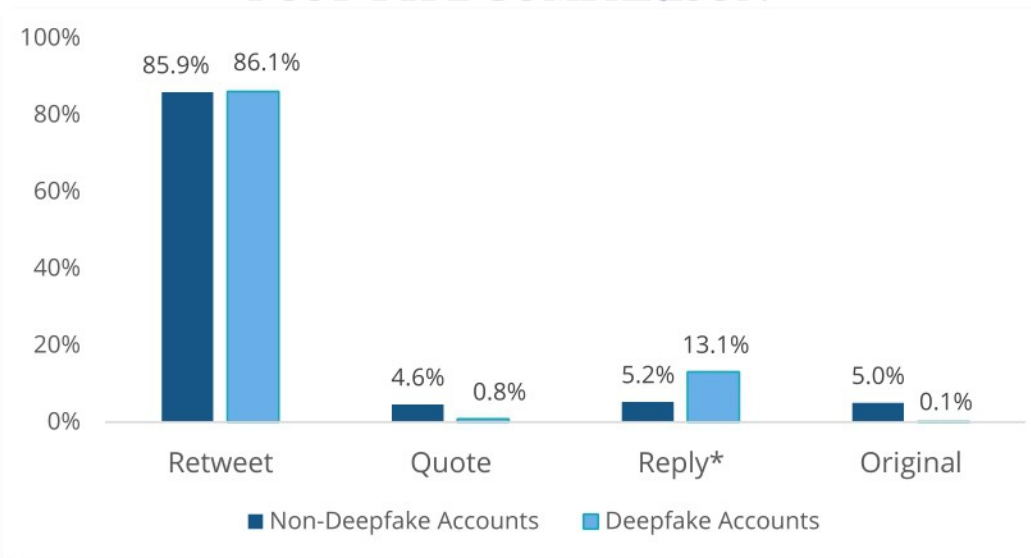
A comparative analysis of deepfake and non-deepfake account engagement with the De Zurda Team and the @cubacooperaven networks uncovered key differences in both network engagement and type of interaction. For example, accounts with deepfake profile images were more active in the De Zurda network, and they posted replies at a greater statistically significant rate. We believe this is further confirmation of substantial human involvement in the administration of these accounts, in part because replies tend to require greater editorial intervention and oversight.

- We compared deepfake account activity to that of non-deepfake accounts, focusing specifically on engagement with @cubacooperaven and/or De Zurda Team accounts (a total of more than 1.84 million tweets posted between 1 June and 17 August).
- Our analysis shows that 89% of the activity for accounts with deepfake profile images was concentrated in the De Zurda Team network, while accounts without deepfake profile images engaged more with @cubacooperaven.

	Non-Deepfake Accounts	Deepfake Accounts
Tweets Engaging with De Zurda Team Accounts	32.1%	89.0%
Tweets Engaging with @cubacooperaven	70.1%	12.0%

Categories not mutually exclusive as a single tweet can engage with both networks.

POST TYPE COMPARISON



* Statistically significant difference ($p < 0.001$)

- Accounts with both deepfake and non-deepfake profile images favored retweets (a common amplification tactic), but the accounts with deepfake profile images posted replies at a statistically significant higher rate. This difference strengthens our assessment that the accounts with deepfake profile images in this network are run by real people. Further, only one of the 16 accounts that posted replies displayed signs of partial [automation](#), such as [tweeting headlines of URLs](#) or [repeating replies](#).⁷

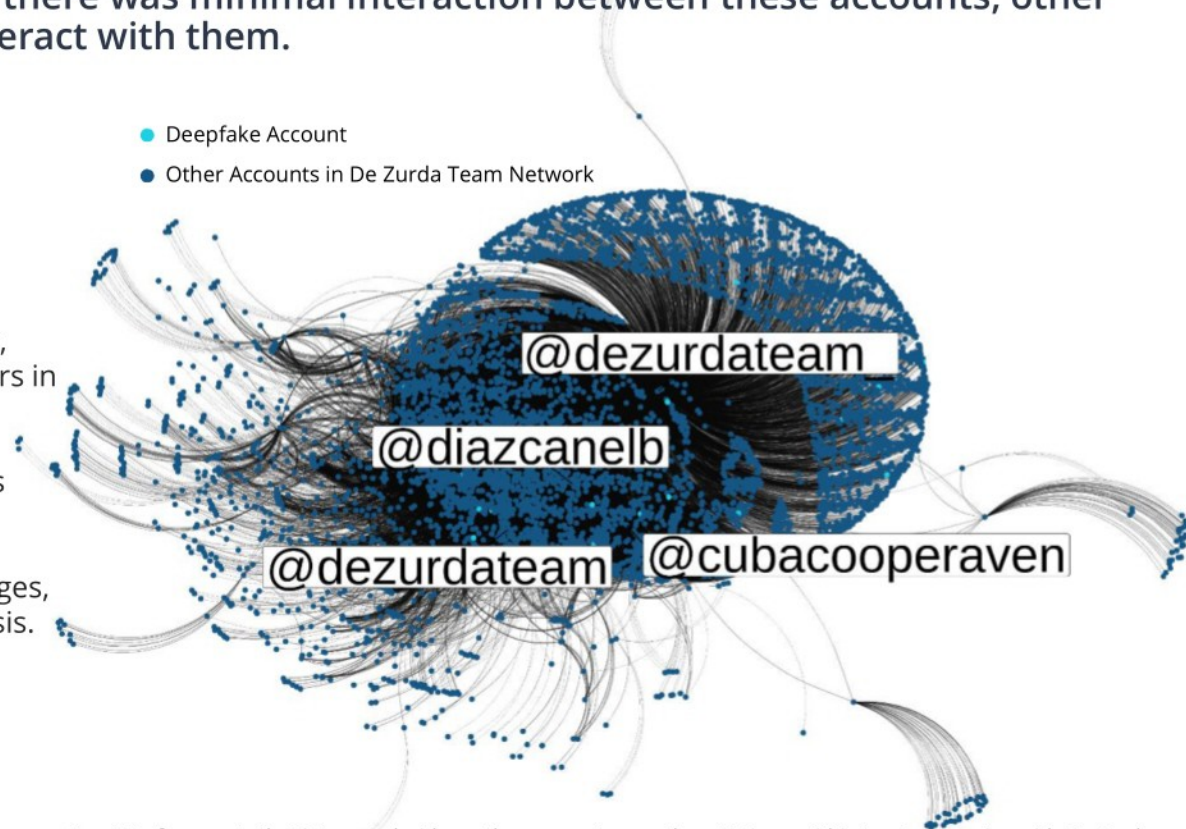
DE ZURDA TEAM NETWORK ANALYSIS

Accounts with deepfake profile images in the De Zurda Team network were spread out, and they actively amplified content from popular Cuban political figures and organizations. While there was minimal interaction between these accounts, other accounts with pro-Cuba images, slogans, and hashtags did interact with them.

- We analyzed mention network activity of engagement with accounts in the De Zurda Team network from between 1 January to 17 August.⁸
- Accounts with deepfake profile images interacted most with those of popular political figures like the president ([@diazcanelb](#)) and vice president ([@drrobertomojeda](#)), and with organizations like the United Left ([@izquierdaunid15](#)) and the Communist Party of Cuba ([@partidopcc](#)).⁹ Notably, these accounts were less likely to interact with the accounts of prominent actors in the network like [@alelross198](#) or [@alirubioglez](#).¹⁰
- The account with a deepfake profile image that garnered the most interactions was [@kami92rey](#), interacting with 401 unique accounts.¹¹
- We observed minimal interaction between accounts with deepfake profile images, with only two pairs of accounts interacting enough to be included in our analysis.

Deepfake Author	Deepfake Interaction Account	Interaction Count
@augustinazahares	@ena_cuba	11
@damianxcuba1	@melye30883863	3
@rubioviamontes	@abiaguiar	1
@kami92rey	@abiaguiar	1
@kami92rey	@maurici77090486	1
@juanfabre1308	@abiaguiar	1

- Deepfake Account
- Other Accounts in De Zurda Team Network



Top 5% of accounts that interacted with another account more than 12 times within tweets engaging with De Zurda Team accounts. Interactions between two accounts is defined as any type of tweet posted by one account that contains a mention of the other accounts (i.e., account A interacts with account B if account A retweets, quote tweets, or replies to tweets posted by account B; if account A retweets another account's post that contains a mention of account B, this would also count as an interaction between accounts A and B).

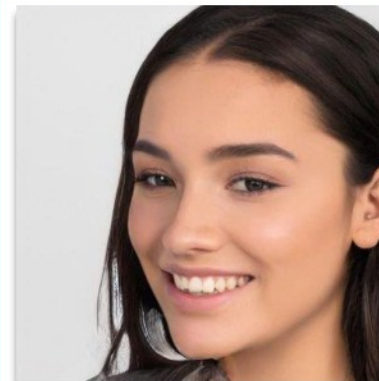
DEEFAKE IDENTIFICATION

Two of the profile images used by accounts in our analysis are ones we have observed previously, and they are similar images that are often reused. While the use of these images may be convincing to a casual social media users, they are easily revealed as inauthentic when subjected to even cursory investigation.

- Two of the deepfake images utilized by accounts in the Cuban networks are popular images encountered by many on the internet, with one of those images figuring prominently in our assessment of a pro-Maduro network.
- The reuse deepfake images typically happens due to the small number of free online providers of deepfake images (e.g., [thispersondoesnotexist](#), [generated.photos](#)) and the high computational burden to create new ones from scratch.¹²



Searching for account [@gadielr13408250's](#) deepfake profile image retrieves the online provider [thispersondoesnotexist](#) and another active [Twitter account](#) using the image.¹³



Account [@jhelenyparede's](#) deepfake image was previously reported by GEC as used by [@aidenevas](#), an account active in a pro-Maduro network.¹⁴ Both accounts are now suspended.

Selected Faces for the 100K Faces Project
[Browse](#) or [Download all \(zip file\)](#)

Searching the image brings up a [Medium post](#) where the image is used as an example of a downloadable collection of generated faces.¹⁵

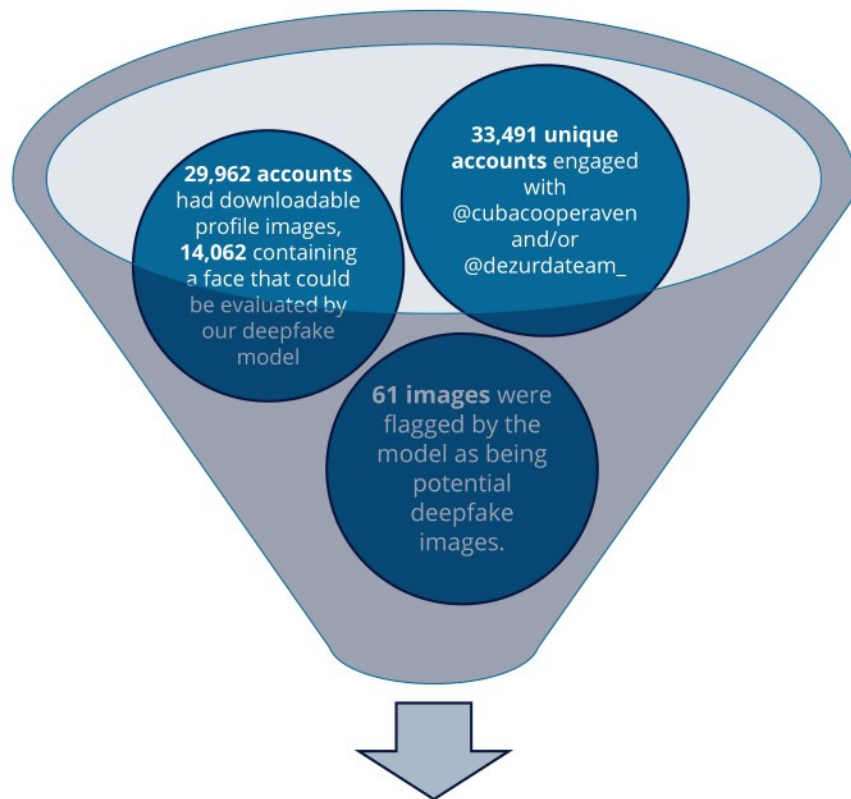


METHODOLOGY AND DEEPPFAKE PROFILE EXAMPLES



DEEPPFAKE IDENTIFICATION METHODOLOGY

Deepfakes are computer generated images that are increasingly used in disinformation operations online. GEC used a custom deepfake detection model to identify 33 accounts active in the @cubacooperaven and/or De Zurda Team networks that used a deepfake profile image.

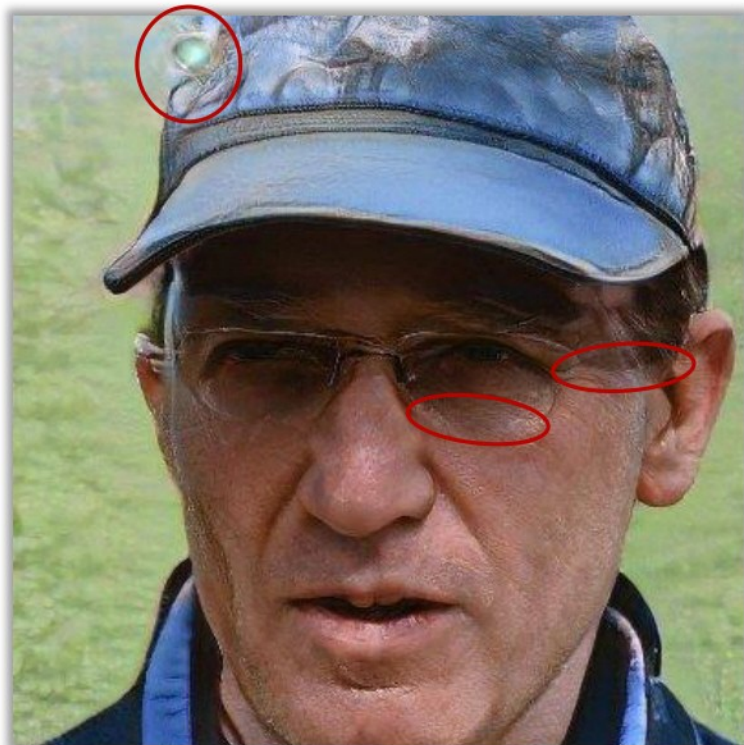


33 accounts manually verified as using a deepfake profile image

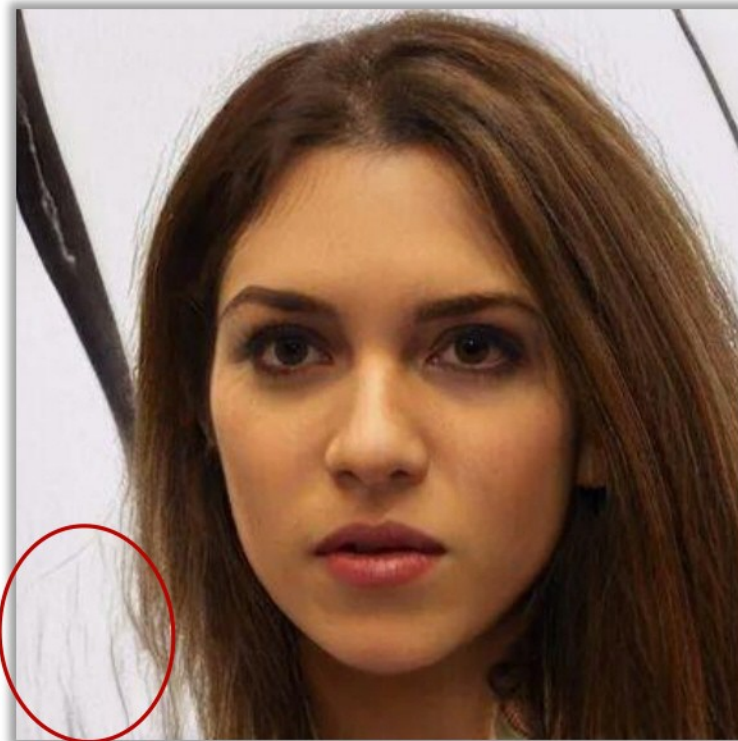
- "Deepfakes" refer to images and videos generated through a generative family of AI models, the most popular of which are generative adversarial networks (GANs).
- Deepfakes have been used in several disinformation campaigns and have been identified as an [emerging threat](#) to spread misinformation, propaganda, and disinformation across social media.¹⁶
- Several disinformation campaigns have used deepfake images as profile photos of [sock-puppet accounts](#).¹⁷ This allows the sock-puppets to appear to be human.
- GEC uses a custom AI algorithm to detect deepfake images. Our detection model can identify images from most popular services providing computer generated faces, but it may fall short with images generated utilizing more sophisticated methods.
 - We applied our deepfake detection model to all available profile images from accounts that interacted with the previous reports focus networks; @cubacooperaven and/or De Zurda Team accounts between 1 June and 17 August.
 - Of 33,491 accounts that engaged with the focus networks, 29,962 had downloadable profile images. Another 14,062 were determined to contain a face by the model's pre-processing and were run through the deepfake detection model.
 - The model predicted that 61 images were likely to be deepfakes. We manually reviewed these and verified that 33 accounts were using a deepfake profile image.

DEEFAKE IDENTIFICATION

Deepfakes are synthetic or manufactured, but extremely realistic, images of people. Such images typically have obscure backgrounds, neutral expressions, and aligned eyes. While facial features can be very convincing, when subjected to closer inspection there are many indicators that an image was GAN-generated. Some examples of these indicators seen in this analysis are included below.



Deepfake image for user [@rubioviamontes](#) contained an artifact, an incohesive area of color or texture, and struggled to create accurate glasses frames.¹⁸



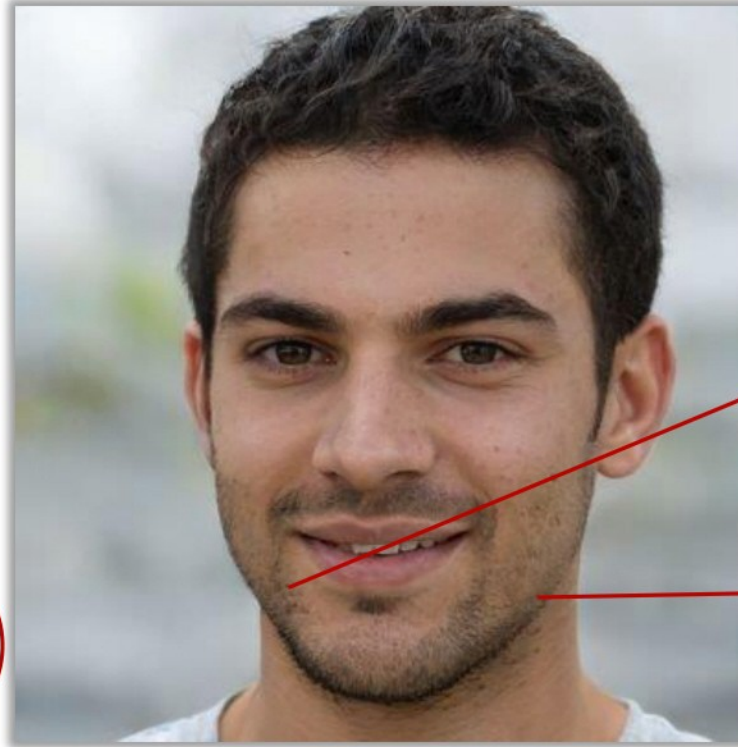
Deepfake image for account [@rubhernandezcar2](#) was identifiable by computer-generated hair not connected to a source.¹⁹



GAN-generated images can struggle with symmetry, sometimes creating mis-matched earrings like in this image for [@aitanaalonso8](#).²⁰



Deepfake image on account [@msoy2020](#) had difficulty creating cohesive clothing.²¹



A more convincing deepfake image for user [@yadielcuba](#) is still identifiable by zooming in on questionable areas.²²



Unrealistic blurry markings and hair angles are indicative of a deepfake.



REFERENCES

1. For more information regarding previous GEC analysis of the @cubacooperaven and De Zurda Team Twitter accounts, please see *"Cuban Government's Propaganda and Disinformation Efforts on Twitter Tied to Pan-Latin America Leftist Support and Cuban Medical Brigades, with Tactics Reminiscent of the Maduro Regime,"* GEC2022-WHA-720, 22 September 2022.
2. In this analysis, GEC identified 25 accounts with deepfake profile images, comprising 0.05% of the 50,000 accounts assessed. For more information regarding deepfake profile images active in the Russia-Ukraine information space, please see *"Use of GAN-Generated Profile Images in the Ukraine-Russia Information Environment,"* GEC2022-GBL-894, 15 March 2022.
3. <https://twitter.com/aitanaalonso8>, <https://twitter.com/gabrielapzerpa>, <https://twitter.com/maurici77090486>, <https://twitter.com/yadielcuba>, <https://twitter.com/rubhernndezcar2>.
4. <https://twitter.com/fuleruso>, <https://twitter.com/tino63154091>, <https://twitter.com/nelsonmaf>.
5. <https://twitter.com/fachucardo64>
6. <https://twitter.com/RicardoPinarRio/status/1488984029026295814>, https://twitter.com/ena_cuba/status/1520712507807219712
7. https://twitter.com/ena_cuba/with_replies, https://twitter.com/ena_cuba/status/1574028926824943617, https://twitter.com/ena_cuba/status/1574016713414279168, https://twitter.com/ena_cuba/status/1574016474649329667
8. We used the 95th percentile as the threshold for consideration in this analysis (over 12 interactions with another account). A total of 18 deepfake accounts surpassed this threshold.
9. <https://twitter.com/diazcanelb>, <https://twitter.com/drrobertomojeda>, <https://twitter.com/izquierdaunid15>
10. <https://twitter.com/alelross198>, <https://twitter.com/alirubioglez>
11. <https://twitter.com/kami92rey>
12. <https://thispersondoesnotexist.com/>, <https://generated.photos/faces>
13. <https://twitter.com/GadielR13408250>, <https://thispersondoesnotexist.com/>, <https://twitter.com/Obestoon>
14. <https://twitter.com/JhelenyParedes>, <https://twitter.com/aidenevas>
15. <https://medium.com/generated-photos/press-aaeb26e632d1>
16. <https://journals.sagepub.com/doi/full/10.1177/2056305120903408>
17. <https://graphika.com/reports/operation-naval-gazing>
18. <https://www.twitter.com/rubioviamontes>
19. <https://twitter.com/rubhernndezcar2>
20. <https://twitter.com/aitanaalonso8>
21. <https://twitter.com/msoy2020>
22. <https://twitter.com/yadielcuba>

Deepfakes Used in Pro-Cuba Networks to Enhance Believability of Disinformation and Propaganda

Deepfake Images Likely Here to Stay

Conclusion

[We would appreciate your feedback by completing a short survey here.](#)

GEC2022-WHA-721



(U) Use of GAN-Generated Profile Images in the Ukraine-Russia Information Environment

Twitter Analysis

Analytics & Research

Global Engagement Center

U.S. Department of State

15 March 2022

GEC2022-GBL-894



(U) EXECUTIVE SUMMARY

(SBU) Of the 50,000+ accounts that published tweets related to the Russian invasion of Ukraine between 9 and 17 February 2022, the GEC identified 25 that used Generative Adversarial Network (GAN) generated images of faces as profile pictures. The GEC further analyzed the most recent tweets published by these 25 accounts and found that five posted pro-Kremlin/anti-Ukraine content, 17 posted anti-Kremlin/pro-Ukraine content, and two posted neutral content.

(U) The pro-Kremlin accounts frequently spread disinformation related to the conflict and criticized Ukraine and the West (especially the United States). Most of these accounts posted original tweets or quote tweets amplifying pro-Kremlin messaging but received little-to-no engagement.

(SBU) The anti-Kremlin accounts criticized Russia for its aggression, its denial of Ukraine's sovereignty, its choice of war over peace, and its use of disinformation. Most of these accounts retweeted popular posts to amplify anti-Kremlin narratives. Original posts from these accounts, however, received little-to-no engagement.

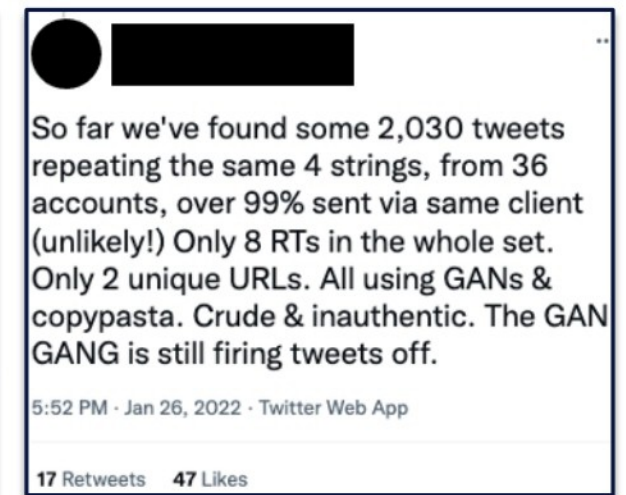
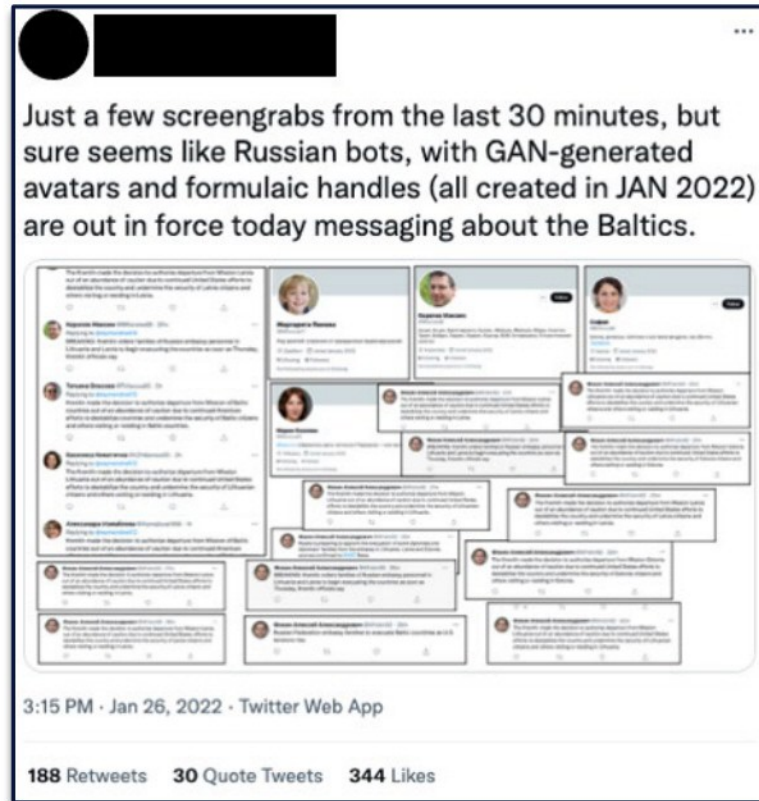
(U) The GEC also identified a large network of likely co-managed accounts that all used GAN-generated profile images and retweeted similar content. These accounts primarily amplified content unrelated to the Russia-Ukraine conflict. The posts that did mention the conflict were largely neutral on the issue.



(U) BACKGROUND

(U) As Kremlin-aligned disinformation efforts continue along with Russia’s invasion of Ukraine, an [emerging Russian tactic](#) appears to involve computer-generated images of human faces, to create the illusion of authentic accounts online. These images are created by the AI technology called Generative Adversarial Network (GAN)—an example of how emerging technology can make inauthentic activity online more difficult to identify. While social media platforms continue to identify and remove accounts with GAN-generated images as profile pictures, the GEC decided to further investigate the existence of these accounts and their influence in the Ukraine-Russia information environment on Twitter.

(SBU)



(SBU)

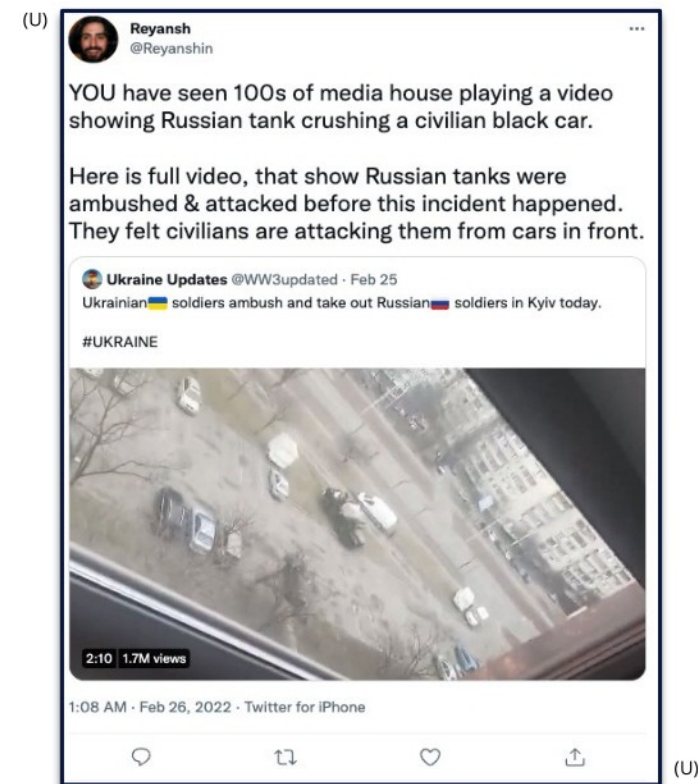


(U) PRO-KREMLIN ACCOUNTS WITH GAN-GENERATED PROFILE PICTURES

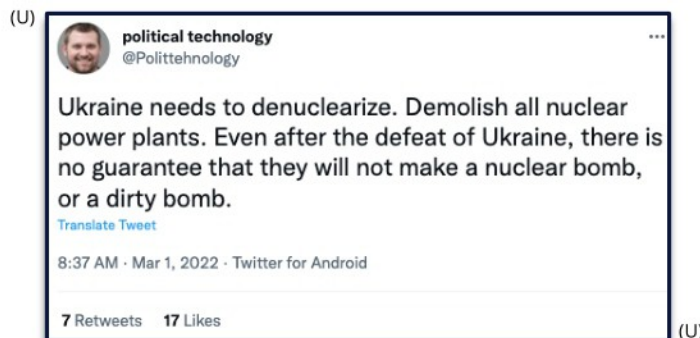
(U) The five identified pro-Kremlin accounts frequently spread disinformation related to the Russia-Ukraine conflict and criticized Ukraine and the West, especially the United States. These accounts posted frequently (two or three posts per day at a minimum), but their posts received little-to-no engagement. These accounts may have used GAN-generated profile pictures to create a sense of trustworthiness—an audience is more likely to trust the legitimacy of tweets from an account with a profile picture of a face than from accounts without a face. GAN-generated images are an example of how emerging technology can make inauthentic activity online more difficult to spot.

(U) Examples of pro-Kremlin narratives and disinformation topics identified:

- Justifications for Russia's invasion of Ukraine, such as the need to [demilitarize](#) or [denuclearize](#) Ukraine.
- Justifications for Russian [attacks on Ukrainian citizens](#) (see example screenshot on the right).
- [Criticism of the West](#), especially the United States (e.g., [blaming the United States](#) for the conflict).
 - This includes negative coverage of U.S. politics—attacks on [President Biden](#) and other [elected officials](#)—as a likely attempt to divert attention from Russian domestic issues.
- Portraying Ukrainian officials as [cowards](#) and spreading content criticizing Ukrainians and [Ukrainian-friendly countries](#).



(U) Example of a tweet supposedly showing a [Ukrainian ambush](#) on Russian soldiers. The tweet attempts to justify why a Russian armored vehicle was seen running over a [civilian vehicle](#).



(U) Example of a tweet demanding the [denuclearization](#) of Ukraine.



(U) Example of a tweet criticizing the [United States and NATO](#) for invading other countries.



F-2023-12348

A-00000828108

"UNCLASSIFIED"

11/18/2024

(U) PRO-KREMLIN ACCOUNTS WITH GAN-GENERATED PROFILE PICTURES

(U) The five pro-Kremlin accounts' GAN-generated profile images are shown below. Of note, @Polittehnology (in the center) was identified in a previous GEC report on pro-Kremlin accounts that amplified negative content about Alexey Navalny and Ukraine.* Additionally, @lexarwabnob's bio contains the text "thispersondoesnotexist," likely a reference to the website thispersondoesnotexist.com, which displays GAN-generated images of human faces.



(U) Left to right: @DedRonin, @Reyanshin, @Polittehnology, @lexarwabnob, @Fixedbets666



(U) GAN-generated faces have numerous artifacts and features that can be used for visual detection. Specifically, GAN-generated faces have centrally located facial features (e.g., eye location is always the same) due to preprocessing techniques required to train the deep learning models that generate these images. If we stack the images on top of each other as shown on the left, we can clearly see that the location of the eyes overlap.



(U) Most GAN-generated images can be easily retrieved from websites such as thispersondoesnotexist.com. Because of limited computational resources, these websites only release a limited number of such images. @Reyanshin's profile picture was actually used in a [Medium article](#) titled "How to recognize fake AI-generated images." The author of the article identifies an example of an artifact in the image (circled in red) that indicates this face is not real.

* Please see "Pro-Kremlin Network Amplifying Negative Content about Navalny and Ukraine Engaged in Inorganic Coordinated Activity with Limited Success," GEC2021-EUR-796, 1 June 2021.



(U) ANTI-KREMLIN ACCOUNTS WITH GAN-GENERATED PROFILE PICTURES

(SBU) Anti-Kremlin accounts criticized both Russia and Vladimir Putin. These accounts often retweeted popular posts, usually those with hundreds or thousands of retweets. However, original posts published by these accounts received little-to-no engagement. Unlike the pro-Kremlin accounts that likely employ GAN-generated images to create an illusion of trustworthiness for their audience, these 17 Russia-based accounts posting anti-Kremlin content possibly use GAN-generated images to protect their identities when criticizing the government. Recently, Russian media regulator [Roskomnadzor](#) announced a ban on any journalist that does not cite "official Russian sources" on the Russia-Ukraine crisis.

(U) Examples of anti-Kremlin narratives identified:

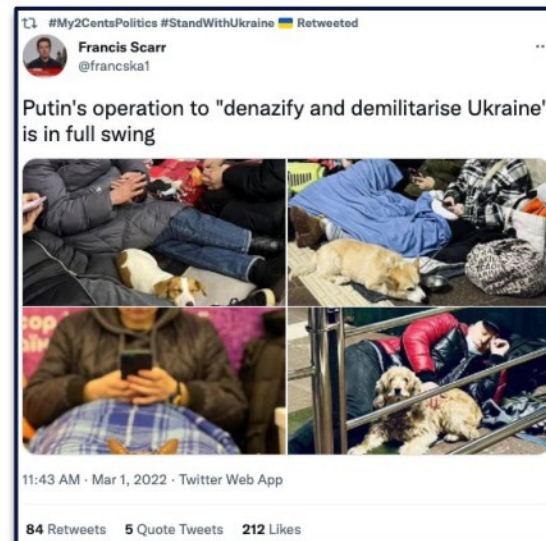
- Criticism of Russia's [aggression](#) and its denial of Ukraine's sovereignty, and its [push for war](#) instead of peace. Notably, all three accounts retweeted the same [post](#) urging Russia to choose a peaceful resolution.
- Criticism of Russia's use of [disinformation](#).
- Portrayals of the difficulties [Ukrainian citizens](#) face due to the ongoing invasion.
- Updates about [Russian attacks](#) in Ukraine.

(U)



(U) An example of an anti-Kremlin account (@GregoryFischer_) that claims to be fake in its bio.

(U)



(U)

(U) A tweet criticizing Putin's justifications for war and portraying the hardships faced by [the affected Ukrainian citizens](#).

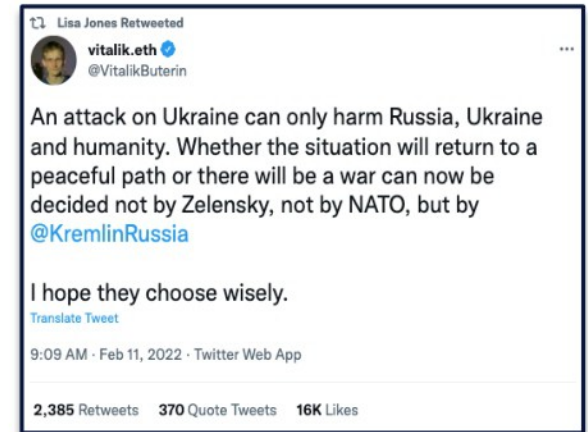
(U)



(U)

(U) Example of a tweet criticizing Russia's use of [disinformation](#).

(U)



(U)

(U) A tweet that was retweeted by three anti-Kremlin accounts urging Russia to choose [peace over war](#).



F-2023-12348 A-00000828108

UNCLASSIFIED

11/16/2024

(U) ACCOUNTS PROMOTING CRYPTOCURRENCY ALSO USE GAN-GENERATED PROFILE PICTURES

(U) Further analysis of the neutral and anti-Kremlin accounts revealed a large network of likely co-managed accounts that all used GAN-generated profile images, co-tweeted/retweeted similar content, and had similar creation dates. Although these accounts did not engage in spreading disinformation nor amplify pro-Kremlin content, it is important to identify and track these types of networks, as they exhibit the same characteristics of communities that often do not engage in spreading disinformation online.

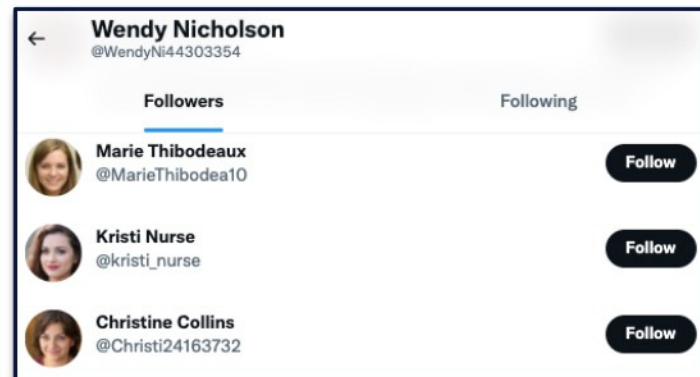
- (U) While these accounts often followed each other, none had more than 60 followers or followed more than 80 accounts. All the identified accounts were created in July 2020.
- (U) All accounts in this network co-tweeted the exact same cryptocurrency-related content. Many of these accounts also retweeted the same posts related to the invasion that were published by the independent Russian media outlet @tvrain (account of TV channel Дождь/Dozhd—also known as *TV Rain*) and the business-focused Russian media outlet @ru_rbc (account of РБК/RBK).
- (SBU) Most of the relevant posts from the accounts in this network took a neutral stance on the conflict (some, however, seemed more anti-Kremlin). Furthermore, most of the posts related to the invasion were in Russian, which further indicates that this network of accounts are co-managed and potentially based in Ukraine or Russia.

(U)



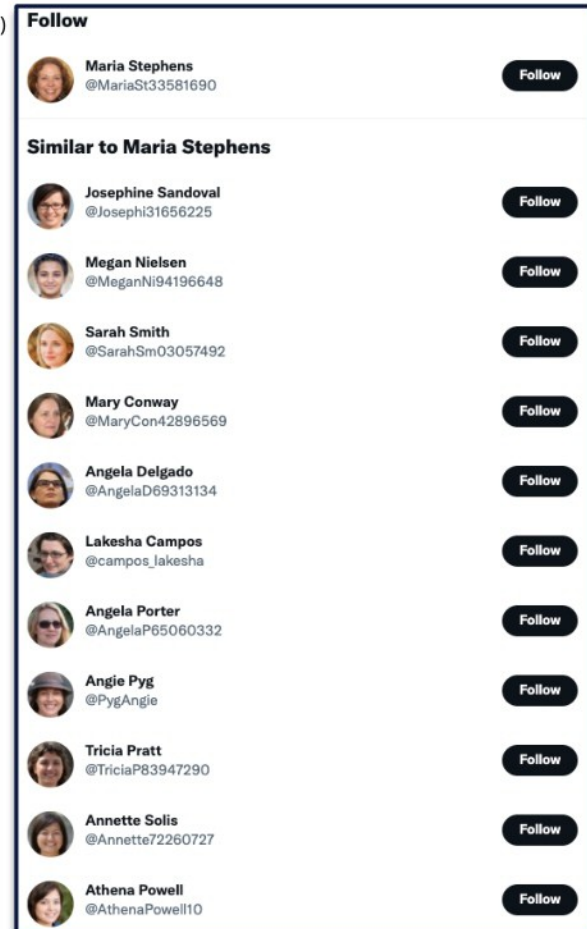
(U) An example of a tweet by @ru_rbc that was retweeted by @MariaSt33581690.

(U)



(U) Example of similar co-managed accounts following @WendyNi44303354.

(U)



(U)

(U) Due to the similar content outputted by these accounts, the Twitter recommendation algorithm identified and recommended similar accounts as @MariaSt33581690. Most (if not all) accounts pictured here use GAN-generated images as their profile picture and are likely part of this network.



(U) REFERENCES

1. <https://about.fb.com/news/2022/02/security-updates-ukraine/>
2. <https://twitter.com/Polittehcnology/status/1497037889506807808>, <https://twitter.com/Polittehcnology/status/1498653582631088128>
3. <https://twitter.com/Reyanshin/status/1497453597457465345>
4. <https://twitter.com/lexarwabnob/status/1498468266808905731>, <https://twitter.com/Fixedbets666/status/1498611676001804289>
5. <https://twitter.com/lexarwabnob/status/1498468742682091521>, <https://twitter.com/lexarwabnob/status/1498794578429845510>
6. <https://twitter.com/DedRonin/status/1498471973856415751>, <https://twitter.com/DedRonin/status/1498089854533525515>
7. <https://twitter.com/Polittehcnology/status/1498653582631088128>
8. <https://twitter.com/Fixedbets666/status/1498693211493179396>
9. <https://twitter.com/Reyanshin/status/1497453597457465345>, <https://www.thedailybeast.com/ukrainian-man-miraculously-survives-russian-tank-running-over-his-car>
10. <https://kcimc.medium.com/how-to-recognize-fake-ai-generated-images-4d1f6f9a2842>
11. <http://washingtonpost.com/technology/2022/02/24/ukraine-russia-war-twitter-social-media/>
12. https://twitter.com/SARFAN_Jan/status/1491443935381581826, https://twitter.com/GregoryFischer_/status/1491829665090125827, <https://twitter.com/LisaJon64108142/status/1492169811295621128>
13. <https://twitter.com/BikePeder/status/1492286734037655557>
14. <https://twitter.com/francska1/status/1498700429701623817>
15. <https://twitter.com/DomCulv/status/1498681294628311047>
16. <https://twitter.com/francska1/status/1498700429701623817>
17. <https://twitter.com/BikePeder/status/1492286734037655557>
18. <https://twitter.com/LisaJon64108142/status/1492169811295621128>
19. https://twitter.com/ru_rbc/status/1496077626787651585



(U) Use of GAN-Generated Profile Images in the Ukraine-Russia Information Environment

Twitter Analysis

CONCLUSION

[We would appreciate your feedback by completing a short survey here.](#)

GEC2022-GBL-894

COORDINATED COMMUNITY & DEEPPFAKE DETECTION

Pro-Kremlin Network Amplifying Negative Content about Navalny and Ukraine Likely Engaged in Inorganic Coordinated Activity with Limited Success

EXECUTIVE SUMMARY: Among a network of pro-Kremlin twitter accounts known to amplify anti-Navalny content, three clusters of accounts engaged in tactics, techniques, and procedures (TTPs) indicative of attempts to coordinate posts and inorganically amplify tweets. One cluster engaged in direct quoting and posted pro-Kremlin content—including posts that were critical of Navalny, the Ukrainian government, and the United States—which, we assess, is a likely indication of inorganic coordinated behavior. Furthermore, one of the accounts in this cluster, @RadioStydoba, is a troll account targeting the Russian-language Radio Free Europe account, @SvobodaRadio. However, the tweets posted by the handles in this cluster received minimal engagement, and thus, their reach was likely limited.

Another cluster of accounts engaged in co-tweeting (posting identical content), which can also be indicative of artificial and coordinated behavior. This tightly intertwined cluster of accounts posted pro-Kremlin content, including negative posts about Navalny and Ukraine. Thirdly, we identified several clusters of highly central accounts that participated in mass retweeting. The two most retweeted accounts in these clusters were @RadioStydoba and @spacelordrock. The latter is formally associated with *RIA-FAN*, which has been sanctioned by the U.S. Department of the Treasury. The most retweeted tweets were negative tweets about Aleksey Navalny and Yulia Navalnaya as well as the Ukrainian government.

While retweets in and of themselves are not explicit signs of coordination, the high volume of tweets, in conjunction with other TTPs, is likely indicative of coordinated inauthentic behavior. Moreover, this was the most effective TTP identified, as measured by engagement and likely reach. Lastly, we examined the profile pictures of accounts in this network and identified several deepfakes using an AI-based detector, which were then manually verified. While deepfakes can be used to maintain anonymity online (which is not inherently nefarious), because some of these accounts amplified pro-Kremlin tweets, they likely also engaged in inauthentic behavior.

REPORT







Coordinated Community Detection

Following a recent network analysis of Twitter accounts, which identified a pro-Kremlin network that amplified anti-Navalny content,¹ the GEC analyzed a sample of 443,747 tweets to identify potential coordination and inorganic amplification, with a focus on the TTPs utilized. These tweets were posted between 1 December 2020 and 13 April 2021 by 11 seed accounts² and 55 pro-regime accounts³ that interacted with the seed accounts mentioned in the previous report, as well as by other accounts that interacted (retweeted, mentioned, or quoted) with these 66 total accounts.⁴ We then specifically examined tweets about Aleksey Navalny and/or Ukraine. The analysis revealed several clusters of accounts, each using slightly different TTPs to amplify either anti-Navalny narratives or narratives supporting Russian interests in Ukraine.

Direct Quoting Cluster

In a cluster containing four accounts, we observed direct quoting—a common coordinated technique to amplify messaging. When an account direct quotes a second account, it copies the content of a tweet of the second account and includes a mention of the second account in the same tweet. The accounts @sornadejda53, @kremlinbolt, and @KarbofosnyjDyh directly quoted @RadioStydoba,⁵ one of the central accounts described in the previous report, at least two to five times each (see image below for example).⁶ While these numbers are low, we assess that these accounts likely engaged inorganic coordination. @RadioStydoba, whose name “Radio Shame” (Радио Стыдоба in Russian) is a word play on the *Radio Free Europe* outlet in Russian (Радио Свобода), has 43,500 followers. Its bio includes links to a Telegram channel with 11,432 subscribers and a YouTube channel with 2,250 subscribers.⁷ The most popular video on the YouTube channel, which has 3,500+ views and was published one day after the pro-Navalny protests on 23 January 2021, urges Russians to respect the riot police.⁸

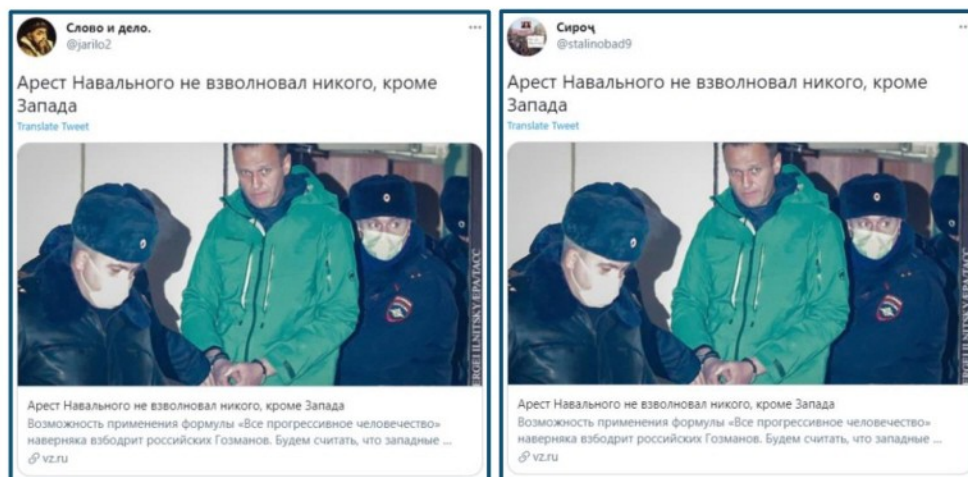
In general, these accounts displayed pro-Kremlin behavior, often tweeting or retweeting posts that criticized and/or ridiculed Navalny, the Ukrainian government, and the United States.⁹ Of note, @KarbofosnyjDyh also amplified misleading content about COVID-19 vaccines, specifically AstraZeneca.¹⁰ This account also posted tweets with links to livejournal.com posts by an account named and_ekb.¹¹ Furthermore, @KarbofosnyjDyh and @kremlinbolt utilized mass retweeting tactics to amplify narratives, whereas @sornadejda53 posted more original tweets. While this small cluster had a minimal impact on the conversation, this tactic was likely used more broadly to amplify content.

 <p>Радио Стыдоба @RadioStydoba</p> <p>На Украине участники митинга против высоких тарифов ворвались в здание областного совета в Житомире, пишут украинские СМИ. Надо было в 2014-м не давать врываться в правительственные здания всяким майдаунам. А теперь поздно, теперь вы - внутренние террористы.</p> <p>Translate Tweet</p> <p>9:49 AM · Jan 15, 2021 · Twitter for iPhone</p> <p>108 Retweets 4 Quote Tweets 675 Likes</p>	 <p>Радио Стыдоба @RadioStydoba</p> <p>В британском посольстве назвали "обычной практикой" приезд дипломатов в суд по Навальному. Кажется высылка дипломатов за вмешательство во внутренние дела страны пребывания - ещё более обычная практика.</p> <p>Translate Tweet</p> <p>5:43 AM · Feb 2, 2021 · Twitter for iPhone</p> <p>105 Retweets 2 Quote Tweets 725 Likes</p>	 <p>Радио Стыдоба @RadioStydoba</p> <p>На Украине уже вторая смерть за сегодня у привитого Астрозенкой.</p> <p>Утром сообщили о смерти военнослужащей в Одесской области, а теперь умер провизор военной аптеки в Черновицкой области. Инсульт. Он привился 15 марта.</p> <p>Translate Tweet</p> <p>1:35 PM · Mar 24, 2021 · Twitter Web App</p> <p>133 Retweets 5 Quote Tweets 325 Likes</p>
 <p>Nadejda S @sornadejda53</p> <p>На Украине участники митинга против высоких тарифов ворвались в здание областного совета в Житомире, пишут украинские СМИ. Надо было в 2014-м не давать врываться в правительственные здания всяким майдаунам. А теперь поздно, теперь вы - внутренние террористы.</p> <p>@RadioStydoba Translate Tweet</p> <p>12:34 PM · Jan 15, 2021 · Twitter for Android</p> <p>1 Like</p>	 <p>pretorianetz 2.0 Вы ещё,казлы,за Белград ответите! @kremlinbolt</p> <p>В британском посольстве назвали "обычной практикой" приезд дипломатов в суд по Навальному. Кажется высылка дипломатов за вмешательство во внутренние дела страны пребывания - ещё более обычная практика.</p> <p>@RadioStydoba Translate Tweet</p> <p>5:52 AM · Feb 2, 2021 · Twitter for Android</p> <p>3 Likes</p>	 <p>ВаларМоргулис @KarbofosnyjDyh</p> <p>На Украине уже вторая смерть за сегодня у привитого Астрозенкой.</p> <p>Утром сообщили о смерти военнослужащей в Одесской области, а теперь умер провизор военной аптеки в Черновицкой области. Инсульт. Он привился 15 марта.</p> <p>@RadioStydoba Translate Tweet</p> <p>1:50 PM · Mar 24, 2021 · Twitter for Android</p> <p>1 Retweet 1 Like</p>

Examples of @sornadejda53, @kremlinbolt, and @KarbofosnyjDyh direct quoting @RadioStydoba. In the first row are original tweets by @RadioStydoba. The second row depicts @sornadejda53, @kremlinbolt, and @KarbofosnyjDyh directly copying @RadioStydoba's content into a new tweet and also tagging @RadioStydoba.

Co-Tweeting Cluster

We also found evidence of coordination from a set of tightly connected accounts that engaged in co-tweeting (posting identical content in a new original tweet), in addition to retweeting, replying, and commenting on each other's posts.¹² @Jarilo2, which has 13,500 followers and links to VK, Telegram, and a blogging site in its bio, was the most active in this set of accounts. The content amplified by this cluster tended to be anti-Navalny or aligned to the Russian regime's interests in Ukraine; an example can be seen below.¹³ We found at least seven instances in which @Jarilo2 co-tweeted other accounts. However, this is likely not inclusive of all the account's activity. In addition to co-tweeting, this network actively replied to or retweeted content from other accounts within the network.¹⁴



Example of @jarilo2 co-tweeting.

Mass Retweeting Clusters

We also identified several clusters of accounts for which tweets from a small number of central accounts were heavily amplified through retweets. The two most retweeted accounts were @spacelordrock and @RadioStydoeba—accounts also previously examined in the GEC's Navalny report series.¹⁵ @spacelordrock, or "Voice of Mordor," is listed on RIA-FAN's website as an associated blogger.¹⁶ RIA-FAN has been linked to Yevgeniy Prigozhin and is physically located in the same building previously occupied by the Internet Research Agency.¹⁷ Additionally, RIA-FAN was sanctioned by the U.S. Department of the Treasury as part of sanctions targeting Project Lakhta.¹⁸ Furthermore, Voice of Mordor is an active contributor to *News Front*, which the GEC has previously identified as a Russian proxy site that produces and perpetuates disinformation. The U.S. Department of the Treasury identified *NewsFront* as an outlet that has worked with the FSB to distribute disinformation and propaganda.¹⁹

This cluster's most retweeted posts, which had hundreds to thousands of retweets, were critical of Aleksey Navalny and his wife Yulia Navalnaya, as well as the Ukrainian government.²⁰ Retweets alone are not inherent signs of coordination; however, because of their high number, they were effective in amplifying and spreading the reach of these posts.

Deepfake Detection

We also analyzed the profile pictures of the 8,063 unique accounts in our dataset. We found several accounts we suspect of using synthetic pictures of people's faces generated with deepfake technology. Synthetically generated deepfaked images are easily accessible to the general public, and they are commonly used by automated and coordinated accounts, or by accounts attempting to anonymize their identities.²¹ We initially identified these synthetic profile pictures using an AI-based detector, which we then manually verified. These types of deepfakes can usually be recognized through artifacts leftover in the image during generation, such as asymmetry (e.g., in jewelry or clothing),²² centrally aligned eyes (due to preprocessing done when training AI models capable of generating deepfakes),²³ and/or obscure/surreal or bland/solid backgrounds.²⁴

@CounterwaveRu2's profile photo is one of the most obvious examples of a synthetic image. This account claims to be a Kremlin-affiliated Russian analyst, and it amplifies pro-Kremlin accounts



@CounterwaveRu2's profile picture and Twitter account page

along with the accounts of Russian embassies around the world. The artifacts are most visible on the left side of the glasses. @Polittehology's profile picture (pictured below) is another suspected deepfake image. The eyes are aligned in the center and the earlobes seem asymmetrical. The background in this image is also almost identical to that of @CounterwaveRu2. Instead of amplifying pro-Kremlin tweets primarily through retweets, this account both posted a high volume of original tweets and opinionated quote tweets (adding original content to retweets).

We also suspect that @mobilekid_ru, @GaryPrickle, and @688hltHEd9DU0dY use deepfake profile images. The profile pictures of @mobilekid_ru and @GaryPrickle contain blurred, surreal backgrounds, which are often characteristics of synthetic images. @688hltHEd9DU0dY's profile photo could potentially be a heavily photoshopped image of a deepfake. These accounts also primarily posted pro-Kremlin content, with retweeting as their main form of amplification. Of note, the deepfakes identified here are likely not all that are present within the network. Rather these are illustrative examples of how this method has been deployed to help amplify narratives.



Profile pictures of @Polittehology, @mobilekid_ru, @GaryPrickle, and @688hltHEd9DU0dY

ANALYST COMMENT: None.

[We would appreciate your feedback by completing a short survey here.](#)

References

¹ For our previous analysis describing the pro-Kremlin network that amplified anti-Navalny content, please see "Pro-Kremlin Regime Twitter Network Amplifies Anti-Navalny Content from Official and Non-State Affiliated Accounts," GEC2021-EUR-381, 29 April 2021.

² The 11 seed accounts were non-state-affiliated Russian accounts that frequently criticized Navalny and whose posts received a relatively high level of engagement.

³ The 55 pro-regime accounts were accounts that amplified content favorable to the Russian government daily at high rates and frequently interacted with the 11 seed accounts through retweets, mentions, etc.

⁴

11 Seed Accounts	55 Interacting Accounts				
@Vityzeva	@1102margomarg o	@bondyрева66	@kotoff7	@popsvami	@sunnipulse786
@27khv	@57_dreem	@cfowymqpnrv2o s	@larissa0406	@publikatsii_ru	@sven40474141
@Grantodov	@81aad8ec4cc043 3	@cvetlanaorlack	@ledidi1012	@r1tilasmi3wiy7x	@sza7uiuinch9djp
@nadiashar	@8qj0r69useaknzs	@dom_a2012	@lenaorexova	@rbicb	@tarikcyrilamar
@indeec1937	@abejlbvictor	@edwolkov	@lifegood1955	@romancolonel	@v5vs6dvplx0b9u o
@leon_elk	@aleksandrkyuch	@efimov2010_cool	@maroslyakova	@ru_bykov	@vedmidek
@pass4twitt	@alkrat78	@gene_gir	@mip1963	@ruskiybelarus	@vitalij11133014
@kompolk	@alla6591	@jarilo2	@notfarmerwife	@samsonovaxeni a	@vn9ihbucf7dbnz p
@spacelordrock	@alleksei2	@jevb8ajozbaeeco	@nsiktr	@slavam1767	@yis85guiwfgztbo
@UrgvRazvedka	@arsmn	@jurik2363	@otec79	@stalinobad9	@zayaz69rus
@RadioStydob a	@avoronina1960	@kebspyncigstzx	@palchun	@stefani580155	@znesd

⁵ <https://twitter.com/RadioStydob>

⁶ @sornadejda53, @kremlinbolt, and @KarbofosnyjDyh directly quoted @RadioStydob at least five, three, and two times, respectively.

⁷ <https://t.me/RadioStydob>; <https://www.youtube.com/channel/UCjxQGCxHXxWGTmavurivW9w>

⁸ https://www.youtube.com/watch?v=UfBirlFi_g0&ab_channel=%D0%90%D0%BB%D0%B5%D0%BA%D1%81%D0%B0%D0%BD%D0%B4%D1%80%D0%A1%D0%B0%D0%BC%D0%BE%D1%85%D0%B2%D0%B0%D0%BB%D0%BE%D0%B2

⁹ <https://twitter.com/kremlinbolt/status/1386916300820262912>;

<https://twitter.com/sornadejda53/status/1387367144761999360>;

<https://twitter.com/KarbofosnyjDyh/status/1370369638526246913>

¹⁰ <https://twitter.com/KarbofosnyjDyh/status/1374780777217544193>

¹¹ <https://twitter.com/KarbofosnyjDyh/status/1374780777217544193>; <https://and-ekb.livejournal.com/profile>

¹² The accounts involved are composed of (and potentially not limited to) @jarilo2, @EdWolkov, @Vityzeva, @stefani580155, @spacelordrock, @stalinobad9, and @stas_stanyslavs.

¹³ <https://twitter.com/jarilo2/status/1351674648786821122>; <https://twitter.com/stalinobad9/status/1351474994795204608>; <https://twitter.com/jarilo2/status/1378297081081823234>; <https://twitter.com/Vityzeva/status/1378280180175736838>

¹⁴ https://twitter.com/stas_stanyslavs/status/1363139029898829825

¹⁵ Other accounts whose accounts were also amplified include @Jarilo2, @UrgvRazvedka, @leon_elk, @kompolk, @indeec1937, @27khv, and @Vityzeva.

¹⁶ <https://twitter.com/spacelordrock>

¹⁷ <https://www.bellingcat.com/news/africa/2020/08/20/threats-lies-and-videotape-prigozhins-long-running-war-on-free-media/>

<https://home.treasury.gov/news/press-releases/sm1118><https://home.treasury.gov/news/press-releases/sm0312>

¹⁸ RIA-FAN, or the Federal News Agency

<https://home.treasury.gov/news/press-releases/sm577>

¹⁹ <https://home.treasury.gov/news/press-releases/jy0126>

<https://home.treasury.gov/policy-issues/financial-sanctions/recent-actions/20210415>

²⁰ <https://twitter.com/radiostydoaba/status/1359906753685692417>;

<https://twitter.com/spacelordrock/status/1341685550600900608>

²¹ Via websites such as <https://thispersondoesnotexist.com/> and <https://generated.photos/>.

²² GANs could have a difficult time generating long-distance dependencies in images of human faces, such as those found between a pair of earrings, the edges of glasses, shapes of ears, and eye color. Therefore, we sometimes see asymmetry in these features.

²³ Before training a GAN to generate synthetic faces, a dataset of real human faces used for the training has to be preprocessed. One of the preprocessing steps involves cropping the images and centering each image around the face. This step also leads to the GAN being able to generate images that are similarly centered around the face in which the position of the eyes in the generated images do not drastically vary.

²⁴ Although the features found on the faces in the training images used for GANs are the same (i.e., all training images have a face with eyes, nose, mouth, etc.), the background in the training images vary. For this reason, GANs generating synthetic faces cannot generate real background scenes unless specifically designed to do so (e.g., by using training images in which the same background is present). Therefore, GANs tend to instead generate background-like-textures which tend to be blurry or unusual.