CS1800 Situation Analysis

Attached is the **Social Media Propaganda** Situation Analysis for CS1800.

Some things to remember:

- **Think multidimensionally**: cyber policy impacts many different issues, and it is important to identify potential risks and opportunities in your analysis. Consider the strengths and weaknesses of several possible responses and select the optimal response.
- **Engage the scenario**: Assume that the situation which we have provided you is plausible. At the same time, think critically about the information that you have been provided and its origins.
- **Consider interests:** Organizations have a broad and diverse set of interests. How might your decision impact other interests which your organization would like to secure? If you choose one course of action, would a different office at your organization reject your approach? Be sure to be able to justify your response as strongly as possible.
- **Think holistically:** It is important to also consider not just your interests, but *all* parties' interests, including states and non-state actors.

What you are responsible for:

1. A brief, informal recorded oral presentation by your group sent to your section TAs via Google Drive. There are no requirements for using visual materials, nor are there requirements for how many members of your group must speak. Expect to outline your course of action for roughly five, but no more than ten, minutes.

2. Responding to at least 3 of the Canvas responses to your presentation.

3. A brief, informal one-page summary of your proposed plan of action to be emailed to your TAs at 11:59 PM the day before your section. Formatting is not especially important – we just want a record of your approach for evaluation purposes. Bullet points are acceptable in this assignment. No bibliography is required.

4. Filling out a peer evaluation after your presentation, during which you will have the opportunity to inform the HTAs about whether all members of your group contributed fairly to your group assignment.

Documents attached:

1. An October 2019 update to Facebook's policy dealing with information operations, entitled "How We Respond to Inauthentic Behavior on Our Platforms: Policy Update"
2. An example of a "takedown document" posted by Facebook for your general reference. This post documented the disruption of two information operations: one in Iraq and another in Ukraine.
3. Facebook's January 2020 policy regarding manipulated media such as deepfakes.

Disclaimer: While some situation analyses for CS1800 are entirely fictional, others are based on real events. You should think about the "date" of your scenario and discount any events that have occurred in the real world after that date.

DATE: May 2, 2020

You are policy analysts in Facebook's Threat Intelligence office who need to make a decision on how to deal with an ongoing, deceptive operation on Facebook and Instagram. Below is an email from your investigations teams:

To: Policy Analysis Team, Threat Intelligence Office, Facebook Inc.
From: Investigations Team, Threat Intelligence Office, Facebook Inc.

Dear Colleagues,

During the course of our regular search for coordinated inauthentic behavior (CIB), we have located an unusual operation currently unfolding on our platforms. It is a network of 130 user accounts, 22 Facebook pages, and 37 Instagram users which are being actively deceptive about their identities and content. The assets were created during a short period in early 2019.

The users purport to be politically right-wing citizens of Adjikistan, a country which has been under a right-wing authoritarian military regime since a violent coup in 2007. The users and pages portray themselves as previous supporters of the military regime, who have recently become disillusioned with its economic policies implemented in 2018. These "average citizens" and pages are now calling for regime change and a return to Adjikistan's pre-2007 democratic constitution. It is worth noting that several of the network's pages have posted an apparent deepfake video purporting to show a senior Adjikistani military general assassinating a political dissident. We have 90% confidence in our assessment that the video is inauthentic.

Our investigation has determined that these accounts are run by a transnational organization called 'Freedom for Adjikistan' (FFA), which is a secretive opposition group comprised of humanitarian activists and aid workers who have sought an end to Adjikistan's authoritarian regime since its establishment in 2011. The network is operated from known FFA technical infrastructure, and much of the accounts are operated from the nearby nation of Libearatova, where many members of FFA operate a government in exile.

We have met with several diplomats at the Department of State to discuss this matter. They have asked us *not* to disrupt or attribute this network publicly. Citing Washington's support for FFA's government in exile and U.S. opposition to Adjikistan's military regime under Generalissimo Adilzhan Monsanto, the Assistant Secretary of State for South and Central Asian Affairs has taken a personal interest in our decision.

However, our policies require us to disrupt and publicly attribute the organizational identity of any actor which is attempting to mislead Facebook users about its identity in order to conduct "information operations," and to act against manipulated media such as deepfakes. Although we have always disrupted such behavior which clearly violates our policies, we are concerned about the major ethical and human rights consequences of disrupting and publicly attributing this campaign. It is worth noting that in 2017, the government of Adjikistan arrested, tried, and executed a seven-member human rights blogging group under the state's terrorism statutes.

Please note that Adjikistan is a major market for our products and services. In 2019, 65% of Adjikistan's 113 million citizens used Facebook, 43% used Instagram, and 77% used WhatsApp. Adjikistan is a middle-income country with a GNI per capita of 11,310 USD and a growing importance in international relations and the global economy. In 2019, Facebook earned 2.1% of its global revenue from Adjikistan. During a 2011 dispute over social media censorship, Generalissimo Monsanto threatened to require his country's internet service providers to block Facebook's services permanently.

We have a commitment as a company to maintain objective policies which are consistently applied. We pride ourselves on judging security violations of users based on their behavior on our platforms rather than the political nature of the content which they post, as long as that content is not calling for violence.

Please respond ASAP with how we should proceed.

Facebook

How We Respond to Inauthentic Behavior on Our Platforms: Policy Update

October 21, 2019

By Nathaniel Gleicher, Head of Cybersecurity Policy

My team coordinates our cross-company effort to find and stop information operations — coordinated campaigns that seek to manipulate public debate — across our platforms. Our coordinated inauthentic behavior (CIB) policy informs how we find, identify and remove these operations. In the past year alone, we have announced and taken down over 50 networks worldwide for engaging in CIB, including ahead of major democratic elections.

As we've improved our ability to disrupt these operations, we have also built a deeper understanding of the different types of threats out there, and how best to counter them. While significant public attention has been on foreign governments engaging in these types of violations, over the past two years, we have also seen non-state actors, domestic groups and commercial companies engaging in this behavior. And we've seen financially-motivated campaigns relying on fake accounts and other inauthentic tactics to drive clicks and mislead people.

While we investigate and enforce against any type of inauthentic behavior, the most appropriate way to address someone boosting the popularity of their posts in their own country may differ from the best way to counter foreign interference. That's why we're updating our inauthentic behavior policy to clarify how we act against the spectrum of deceptive practices we see on our platform. In this post, I'll share more about our thinking and the policies we put in place to tackle information operations.

What is Inauthentic Behavior?

When people think about information operations, they often focus on the content that is being shared. Is it hate speech? Is it a threat? Is it false? But most of the content shared by coordinated manipulation campaigns isn't provably false, and would in fact be acceptable political discourse if shared by authentic audiences. The real issue is that the actors behind these campaigns are using deceptive behaviors to conceal the identity of the organization behind a campaign, make the organization or its activity appear more popular or trustworthy than it is, or evade our enforcement efforts. **That's why, when we take down information operations, we are taking action based on the behavior we see on our platform — not based on who the actors behind it are or what they say.**

Inauthentic Behavior Policy Update

We're updating our inauthentic behavior policy to further improve our ability to counter new tactics and bad actors. Here's how we will act against a range of inauthentic activities, whether foreign or domestic, state or non-state.

Coordinated Inauthentic Behavior

As with our previous takedowns, we will continue looking for groups of accounts and Pages working together to mislead people about who they are and what they're doing. When we find domestic, non-government campaigns in which the use of fake accounts is central to the operation, we will remove all inauthentic and authentic accounts, Pages and groups directly involved in this activity. We will share our findings as part of a regular CIB report which we'll launch in the coming months. However, if the activity we remove is directly related to a civic event, poses imminent harm or involves a new technique or a new significant bad actor, we will share our findings at the time of enforcement.

- **Persona Non Grata**: If, in the course of a CIB investigation, we determine that a particular organization is primarily organized to conduct manipulation campaigns, we will permanently remove it from our platforms in its entirety.

Foreign or Government Interference (FGI)

There are two types of CIB that are particularly egregious:

- Foreign-led efforts to manipulate public debate in another country
- Operations run by a government to target its own citizens. These can be particularly concerning when they combine deceptive techniques with the real-world power of a state.

If we see any instances of CIB conducted on behalf of a government entity or by a foreign actor, we will apply the broadest enforcement measures including the removal of every on-platform property connected to the operation itself and the people and organizations behind it. We will also announce the removal of this activity at the time of enforcement.

Inauthentic Behavior (IB)

We routinely take action against other inauthentic behaviors, including financially-motivated activity like spam or fake engagement tactics that rely on inauthentic amplification or evading enforcement, rather than a core use of fake accounts. We've updated a list of these deceptive techniques in our policy to make sure people better understand the types of inauthentic behavior we will enforce against, even if it's not part of a CIB campaign. We enforce against IB only based on specific protocols that are reviewed and approved through our internal process. It may include temporary restrictions, warnings, down-ranking or removal.

We will continue to adapt our policies to ensure we can effectively combat information operations even as bad actors evolve their techniques.

Facebook

Removing Coordinated Inauthentic Behavior From Iraq and Ukraine

September 16, 2019

By Nathaniel Gleicher, Head of Cybersecurity Policy

Today, we removed multiple Pages, Groups and accounts that were involved in coordinated inauthentic behavior on Facebook and Instagram. We found two separate, unconnected operations that originated in Iraq and Ukraine. We didn't find any links between the campaigns we removed, but they both created networks of accounts to mislead others about who they were and what they were doing.

We're constantly working to detect and stop this type of activity because we don't want our services to be used to manipulate people. We're taking down these Pages, Groups and accounts based on their behavior, not the content they posted. In each of these cases, the people behind this activity coordinated with one another and used fake accounts to misrepresent themselves, and that was the basis for our action. We have shared information about our analysis with our industry partners.

We are making progress rooting out this abuse, but as we've said before, it's an ongoing challenge. We're committed to continually improving to stay ahead. That means building better technology, hiring more people and working more closely with law enforcement, security experts and other companies.

What We've Found So Far

We removed 76 accounts, 120 Facebook Pages, one Group, two Events and seven Instagram accounts for engaging in domestic-focused coordinated inauthentic behavior in Iraq. The people behind this activity used fake accounts to amplify their content and manage Pages — some of which were likely purchased. Many of these Pages merged with one another and changed names over time. They also impersonated other people and used their IDs to conceal their identity and attempt to avoid detection and removal. The Page admins and account owners typically posted about domestic political and societal issues such as religion, various public figures including Saddam Hussein, the state of the military under the Saddam rule, tensions with Iran, the US military action in Iraq, Iranian-backed militia operating in Iraq and Kurdish-Iraqi politics.

- *Presence on Facebook and Instagram:* 76 accounts, 120 Facebook Pages, 1 Group, 2 Events and 7 accounts on Instagram.
- *Followers: Less than* 1.6 million accounts followed one or more of these Pages, about 339,000 accounts joined at least one of these Groups and around 2,000 people followed one or more of these Instagram accounts.
- *Advertising:* Less than $1,600 spent on Facebook and Instagram ads paid for in US dollars, Canadian dollars and Malaysian ringgits.

- *Events:* 2 Events were hosted by these Pages. The first was scheduled for February 2016 and the most recent was scheduled for May 2016. Up to 15 people expressed interest in at least one of these events. We cannot confirm whether any of these events actually occurred.

We identified these accounts through our investigation into suspected coordinated inauthentic behavior in the region.

Below is a sample of the content posted by some of these Pages.

Page name translation: Iraqi in Turkey
Caption: Statement on the 15th anniversary of the first Falluja battle. 4/4/2004. In the Name of Allah the Most Merciful...

Page name translation: Peace Bird **Caption:** I wanted to present to you a bouquet of flowers in celebration of the blessed Eid Al Adha... But we couldn't find better than the smiling faces of men in their most difficult situations... An evidence of their bravery and manhood. Don't choose the leaders that exaggerate their victory and dismiss their responsibility of failure.

Page name translation: Al Abbas News Network
Caption: #BreakingNews Happening Now || Sheikh Faisal al-Issawi and the tribes of Ameriya Fallujah and the security forces with the participation of coalition warplanes launched an attack on the area of Boadij towards the intersection of Al Salam and the area of Hosa, a number of houses in the area have been seized.

Page title translation: Author Raghd AlGabri **Caption:** To those who say that Iraq is the aggressor against Iran, here is a summary. Please engage with this post in all the groups. #Lost_Details_We_Present_To_You #Reasons_We_Are_At_War_With_Iran **Yellow snipe on post:** "Reasons that led us to be at war with evil Iran.. Exclusive to "Bird of Peace" page"

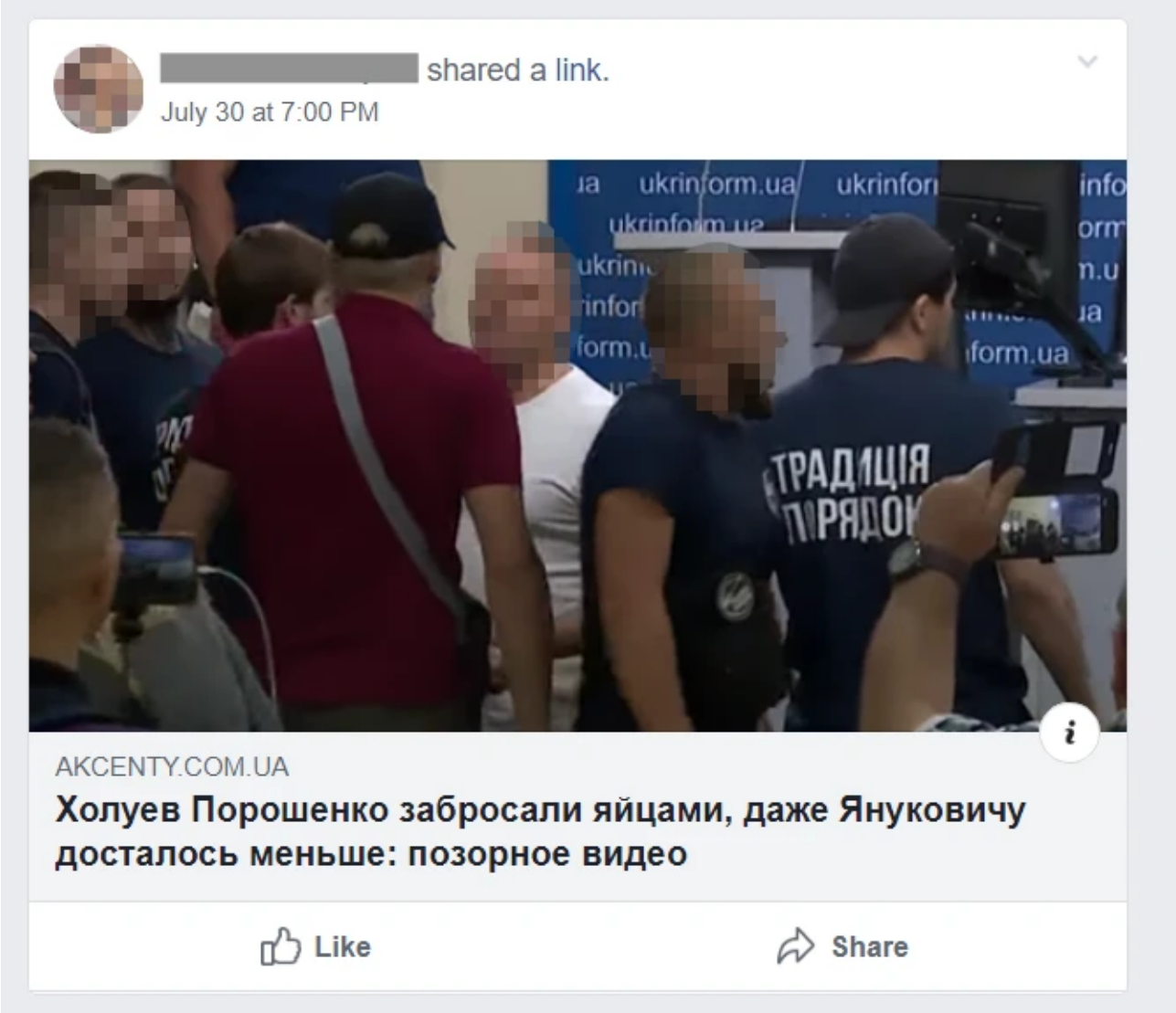
We also removed 168 accounts, 149 Facebook Pages and 79 Groups for engaging in domestic-focused coordinated inauthentic behavior in Ukraine. The people behind this activity used fake accounts to manage Groups and a number of Pages — some of which changed their names over time, and also to increase engagement, disseminate content and drive people to off-platform sites posing as news outlets. The Page admins and account owners typically posted about celebrities, show business, sports, local and international news, political and economic issues including Ukrainian elections, political candidates and criticism of various public figures. Although the people behind this activity attempted to conceal their identities, our review linked this activity to Pragmatico, a Ukrainian PR firm.

- *Presence on Facebook:* 168 Facebook accounts, 149 Pages and 79 Groups.

- *Followers: Less than* 4.2 million accounts followed one or more of these Pages and about 401,000 accounts joined at least one of these Groups.
- *Advertising:* About $1.6 million spent on Facebook and Instagram ads paid for in US dollars.

We identified these accounts through our internal investigation into suspected coordinated inauthentic behavior in the region.

Below is a sample of the content posted by some of these Pages.



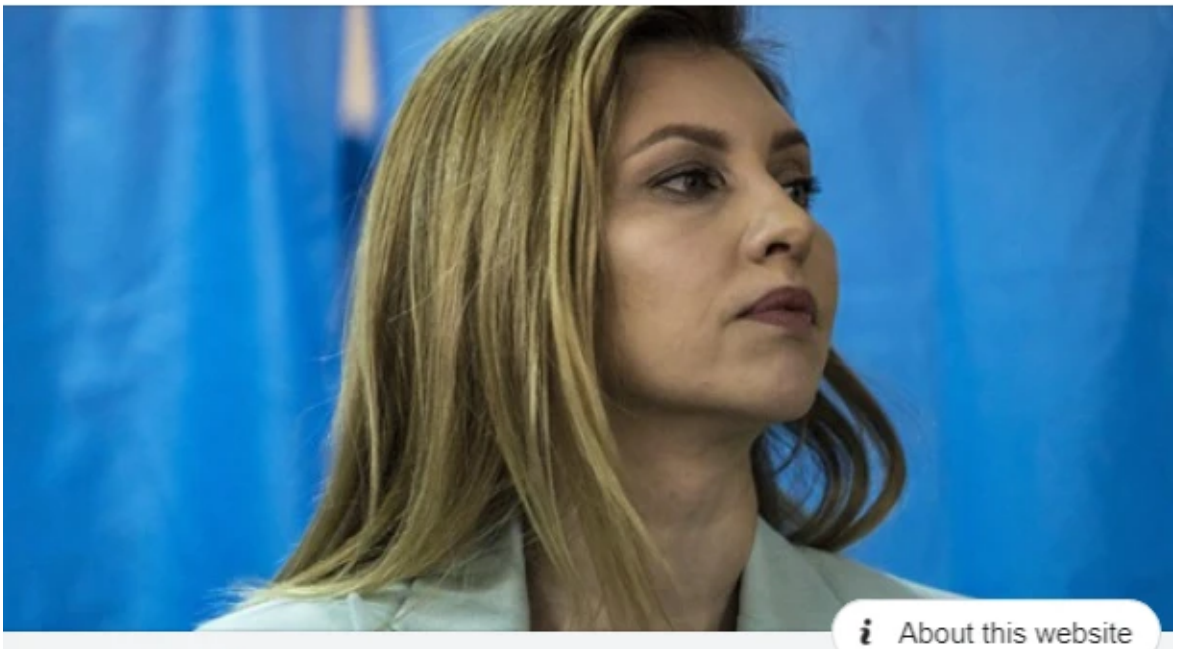
Translation: Poroshenko's minions were egged, even Yanukovich got better treatment: embarrassing video

FACEBOOK

Hyser

а вот такого никто не ожидал!

See Translation



HYSER.COM.UA

Секретное письмо Елене Зеленской просочилось в сеть. Там пишут о ее муже

Caption: No one expected this! **Headline:** A secret letter to Elena Zelenskaya leaked online. It is about her husband

FACEBOOK

Politeka
6 hrs

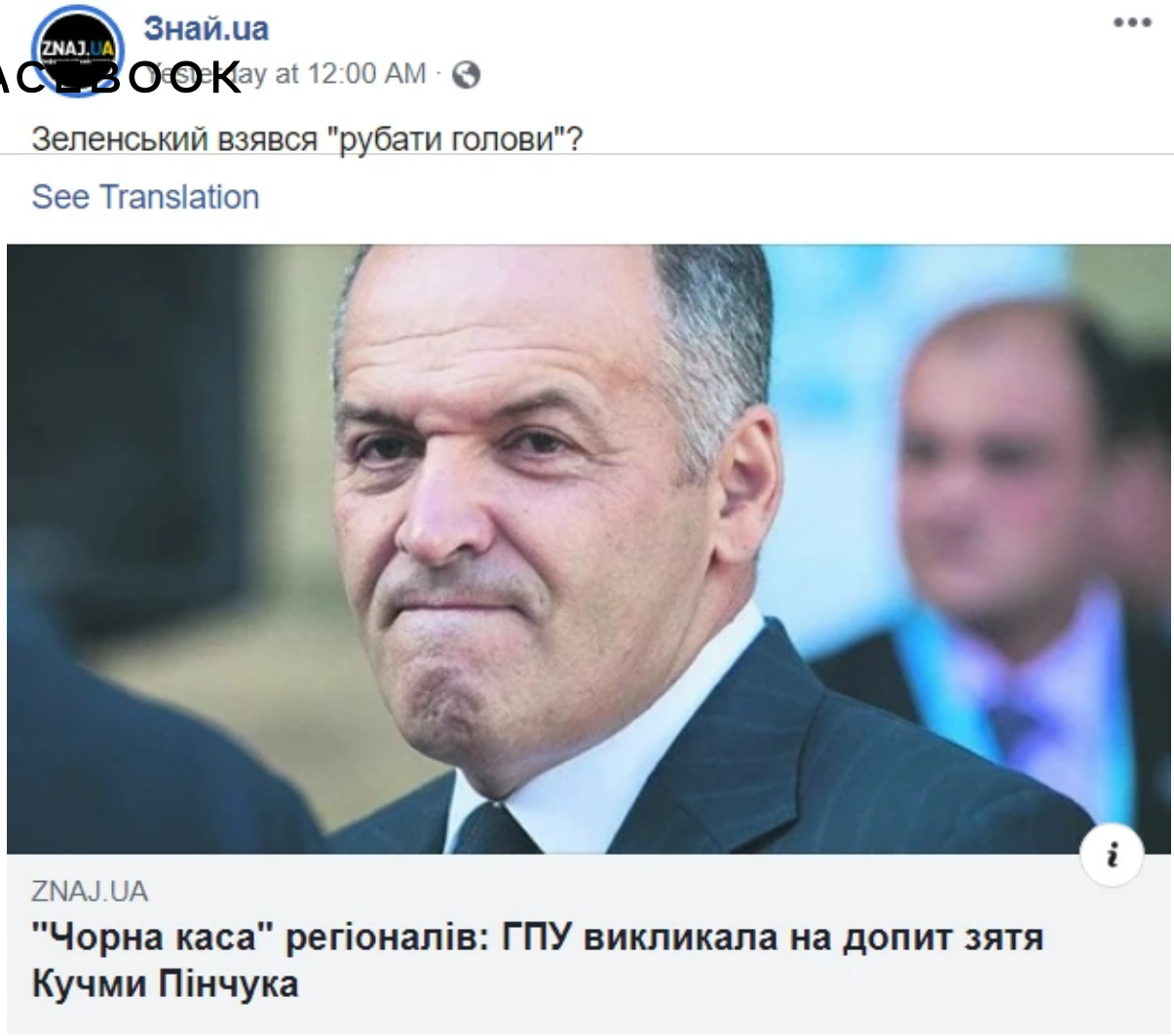
Позор увидела вся Украина!

See Translation

POLITEKA.NET

Геращенко поплатилась за "Зеленского-неуча", позор увидела вся Украина

Caption: The whole of Ukraine saw this embarrassment! **Headline:** Gerashenko was paid back for “illiterate Zelensky”, the embarrassment seen by whole of Ukraine



Caption: Zelenskiy started “cutting off the heads”? **Headline:** “Black cash box” of the regionals: General Prosecutor’s Office summons Kuchma’s son-in-law Pinchuk for questioning

Categories: Facebook, Integrity and Security

Tags: Coordinated Inauthentic Behavior

Related News

Facebook

Enforcing Against Manipulated Media

We’re strengthening our policy toward misleading manipulated videos that have been identified as deepfakes.

January 6, 2020

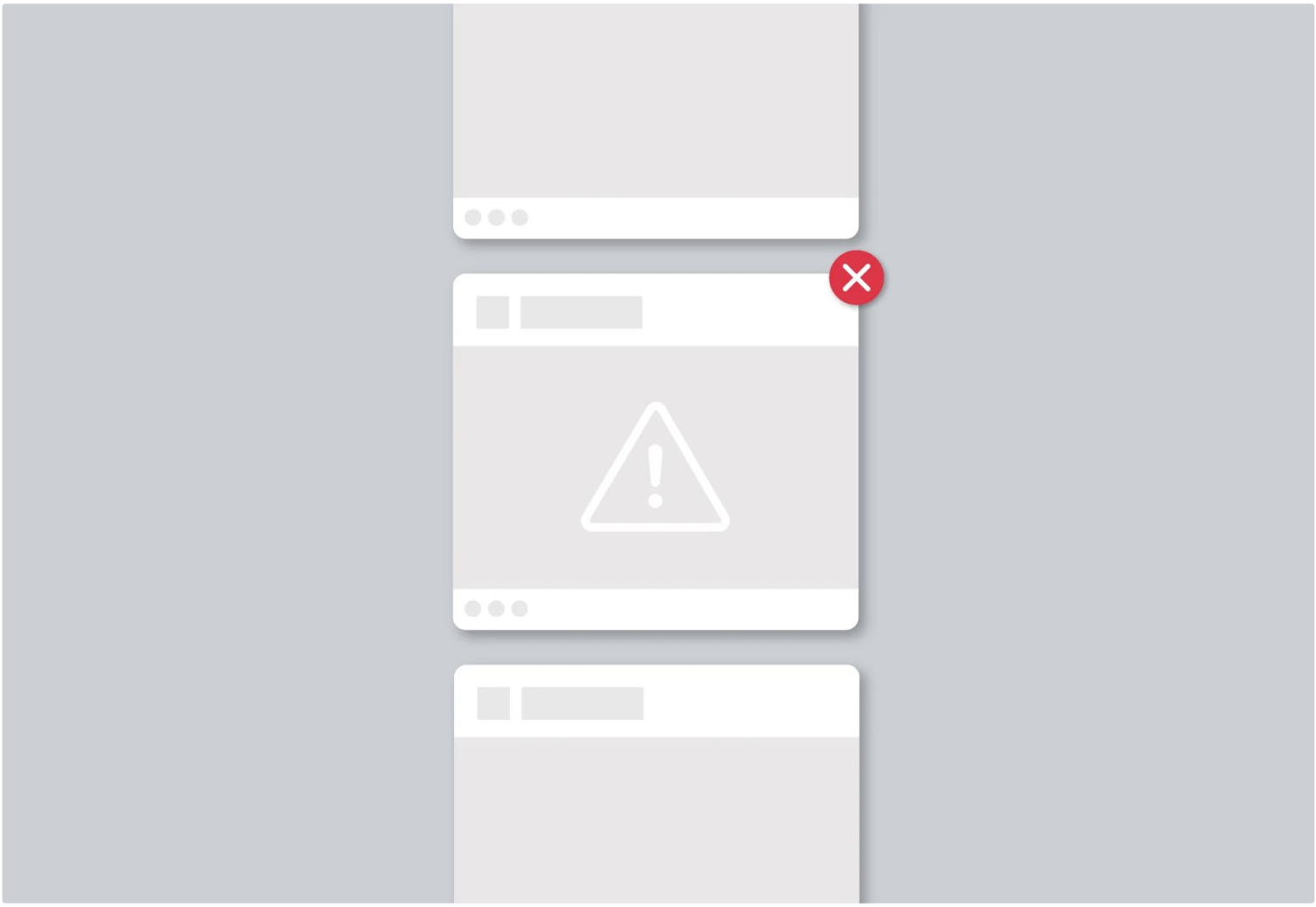
Back to Newsroom

Facebook

Enforcing Against Manipulated Media

January 6, 2020

By Monika Bickert, Vice President, Global Policy Management



People share millions of photos and videos on Facebook every day, creating some of the most engaging and creative visuals on our platform. Some of that content is manipulated, often for benign reasons, like making a video sharper or audio more clear. But there are people who engage in media manipulation in order to mislead.

Manipulations can be made through simple technology like Photoshop or through sophisticated tools that use artificial intelligence or "deep learning" techniques to create videos that distort reality – usually called "deepfakes." While these videos are still rare on the internet, they present a significant challenge for our industry and society as their use increases.

Today we want to describe how we are addressing both deepfakes and all types of manipulated media. Our approach has several components, from investigating AI-generated content and deceptive behaviors like fake accounts, to partnering with academia, government and industry to exposing people behind these efforts.

Collaboration is key. Across the world, we've been driving conversations with more than 50 global experts with technical, policy, media, legal, civic and academic backgrounds to inform our policy development and improve the science of detecting manipulated media.

As a result of these partnerships and discussions, we are strengthening our policy toward misleading manipulated videos that have been identified as deepfakes. Going forward, we will remove misleading manipulated media if it meets the following criteria:

- It has been edited or synthesized – beyond adjustments for clarity or quality – in ways that aren't apparent to an average person and would likely mislead someone into thinking that a subject of the video said words that they did not actually say. And:
- It is the product of artificial intelligence or machine learning that merges, replaces or superimposes content onto a video, making it appear to be authentic.

This policy does not extend to content that is parody or satire, or video that has been edited solely to omit or change the order of words.

Consistent with our existing policies, audio, photos or videos, whether a deepfake or not, will be removed from Facebook if they violate any of our other Community Standards including those governing nudity, graphic violence, voter suppression and hate speech.

Videos that don't meet these standards for removal are still eligible for review by one of our independent third-party fact-checkers, which include over 50 partners worldwide fact-checking in over 40 languages. If a photo or video is rated false or partly false by a fact-

checker, we significantly reduce its distribution in News Feed and reject it if it's being run as an ad. And critically, people who see it, try to share it, or have already shared it, will see warnings alerting them that it's false.

This approach is critical to our strategy and one we heard specifically from our conversations with experts. If we simply removed all manipulated videos flagged by fact-checkers as false, the videos would still be available elsewhere on the internet or social media ecosystem. By leaving them up and labelling them as false, we're providing people with important information and context.

Our enforcement strategy against misleading manipulated media also benefits from our efforts to root out the people behind these efforts. Just last month, we identified and removed a network using AI-generated photos to conceal their fake accounts. Our teams continue to proactively hunt for fake accounts and other coordinated inauthentic behavior.

We are also engaged in the identification of manipulated content, of which deepfakes are the most challenging to detect. That's why last September we launched the Deep Fake Detection Challenge, which has spurred people from all over the world to produce more research and open source tools to detect deepfakes. This project, supported by $10 million in grants, includes a cross-sector coalition of organizations including the Partnership on AI, Cornell Tech, the University of California Berkeley, MIT, WITNESS, Microsoft, the BBC and AWS, among several others in civil society and the technology, media and academic communities.

In a separate effort, we've partnered with Reuters, the world's largest multimedia news provider, to help newsrooms worldwide to identify deepfakes and manipulated media through a free online training course. News organizations increasingly rely on third parties for large volumes of images and video, and identifying manipulated visuals is a significant challenge. This program aims to support newsrooms trying to do this work.

As these partnerships and our own insights evolve, so too will our policies toward manipulated media. In the meantime, we're committed to investing within Facebook and working with other stakeholders in this area to find solutions with real impact.

Categories:
Facebook, Facebook Company, Integrity and Security, Public Policy, Safety and Expression