SafetyComm.
SOUTH SUDAN

211CHECK
Fact-checking and information verification in South Sudan

Date: 24th September, 2024

First: Joint Advisory by SafetyComm and 211CHECK on Social Media Cyber Threats and Scams in South Sudan

Advisory Key Areas:

- Inauthentic false alarms on Facebook page verification
- False alarms on Meta Community standards violations, and
- False alarms on copyright infringement with a threat to shut down an account.

Overview:

This joint advisory by SafetyComm and 211CHECK aims to alert the general public about inauthentic false alarms on Facebook page verification, community standards violations, and copyright infringement with threats to shut down users' accounts. These false alarms are being sent to many users on social media platforms, especially Facebook, by actors claiming to be from Meta or a Facebook company.

These actors target individual profiles, pages, or accounts with malicious notification message campaigns against celebrities in the music industry, digital creators, influencers, public figures, individuals, and accounts of institutions at different levels.

However, it appears that these messages are not directly from Meta but from individuals impersonating Meta or Facebook.

The public is urged to pay attention and avoid clicking links attached to messages they receive, whether through email or direct messages in their social media accounts, before they can verify.

In an unprecedented trend of cyber scams, SafetyComm and 211 Check have seen incidents of social media cyber scams, where some users have been receiving false, inauthentic alarm emails or direct messages on page verification allegedly from Meta.

These users received notifications informing them of their acceptance by Meta to have verified badges on their profiles, claiming they met the Meta badge verification requirements and were eligible to apply within 24 hours through a link to have the blue tick badge on their accounts.

Other users have received notifications through direct messages notifying them that they have violated Meta community standards by using someone else's fake/photo, sharing content that misleads other users, and insulting others. As such, the notification message urged them to click a link within 24 hours to request a review if they thought it was a mistake.

Besides, other users have received notifications alleged to be from the Facebook Support Team that their accounts were scheduled for permanent deletion because of their content, which allegedly has infringed upon Meta trademark rights. In the message, users were asked to request a review through a link; if not, then their accounts would be deleted permanently.

How to detect inauthentic false emails or direct message alarms that claim to be from Facebook or Meta:

To detect if such an alarm notification message is fraudulent or a scam and not from Meta or Facebook, an individual needs to critically analyse the message and the email header or address. What normally appears unusual is that there is an appeal to urgently act to click a malicious link to initiate a process; this is to bait a user to act quickly within the limited time the actors set. In other instances, the email address used to send the message is not official and does not align with the Facebook website domain.

To verify whether an email message is from Facebook, a user needs to log in to his or her personal profile/account but not the page, then go to the account center, and click the recent email button; this will reveal whether there is any new/recent email Facebook sent to the user or not. Legitimate messages from Facebook directly address the users by name or username, however, scam messages use generic greetings like "Dear user" to mask their impersonation. In case of user violation, Facebook will not immediately disable your account, but first warnings and temporary restrictions; any irrational threats indicate a scam. Facebook will never ask for your password or disablement repeal through a link attached to an unsolicited message, this means any request to share user login credentials with a third party is a huge red flag.

Contact Information:

For further assistance or to report a scam, please contact our incident response desk at:

Phone: +211920050106
+211921350435
Email: info@safetycomm.org
info@211check.org

Stay Alert. Stay Protected.
SafetyComm South Sudan and 211 Check

SafetyComm.
SOUTH SUDAN

211CHECK
Fact-checking and information verification in South Sudan

Date: 24th September, 2024

Second: Joint Advisory by SafetyComm and 211 Check on Social Media Cyber Threats and Scams in South Sudan

Advisory Key Area:

- How authentic do the Facebook page verification process, Meta Community Standards violations, and copyright infringement look like?

Overview:

In any violation, Meta allows users to file an appeal directly within the platforms, either Facebook or Instagram. Normally, users are prompted with a "Request Review" button when their content or page is flagged or removed. This option appears on the notification they receive, allowing them to ask Facebook to review the decision.

For business pages, Meta offers the appeal option through the Business Manager. Page administrators can log into their accounts, navigate to the flagged content, and follow the steps provided to appeal the action. Meta emphasises that users communicate directly through official channels, not through Messenger or unsolicited emails. If there is a genuine issue with the user's account or content, the appeal option will appear directly on the platform, ensuring that users can address the concern without needing to click on external links.

How authentic do the Facebook page verification process, Meta Community Standards violations, and copyright infringement look like?

Facebook page verification:

Facebook page verification is a process through which Facebook authenticates the identity of a profile or page to ensure it belongs to a legitimate entity, such as a business, public figure, celebrity, digital creator, or organization. Verified pages display a blue or grey badge next to their name, signalling to users that the page is authentic.

Meta has two types of verifications. First, the verified badge is a criterion for notable persons, brands, or entities that have been featured in multiple news sources but not on paid or promotional content, and then the accounts would be verified by Meta. The second is a Meta Verified as another criterion for a paid subscription that comes with features like a verified badge, account support, and impersonation protection. Literally, these are the requirements or processes, together with following Community Standards, to have a profile or page verified, whether on Facebook or Instagram.

However, only one page or profile per person or business may be verified, with exceptions for language-specific pages and profiles, and general interest pages and profiles can not be verified, for example, Puppy Memes, according to Meta policy.

To get verified on Meta platforms like Facebook or Instagram, a user needs to submit a request through the Facebook Help Center, providing proof of authenticity, such as a government-issued ID or official business documents. The application requires that the account be complete, public, and represent a notable figure, brand, or entity.

To apply for the Meta Verified subscription service, users need to meet eligibility requirements, such as providing a government ID that matches the profile name and photo, and subscribe via the app's settings.

Meta or Facebook does not send soliciting messages attached with links to users via email or direct message to start the verification process.

Meta Community standards violations:

Meta's Community Standards are guidelines designed to foster a safe and respectful online environment across its platforms (Facebook, Instagram, etc.). These standards address content that might harm individuals or communities.

Meta's Community Standards Violations occur when users or pages engage in activities or post content that contravenes the guidelines set to ensure safety, integrity, and respect on platforms like Facebook and Instagram. Here are some common types of violations, along with examples of how users, pages, or accounts breach these policies:

Incitement to Violence

Violations of this nature occur when users post content that encourages or glorifies violence. For example, a post calling for harm against a specific ethnic group or organisation, sharing videos or images glorifying acts of terrorism or promoting criminal behaviour, and posting threats targeting individuals, groups, or public figures, such as encouraging physical assault or harassment. In another instance, if a user posts a video inciting violence during a political protest, encouraging viewers to attack certain individuals. Meta's systems flag this as a violation, potentially leading to account suspension or restriction.

Hate Speech

Hate speech violations arise when content demeans people based on race, religion, ethnicity, gender, or sexual orientation. Meta violations include the use of racial slurs or derogatory terms aimed at dehumanising others, content that denies the occurrence of historical events tied to systemic hatred, such as the Holocaust, and posts that label a group as inferior or subhuman, inciting social hatred. For example, a page that posts content that mocks individuals based on their disability or sexual orientation violates standards on hate speech and will lead to the removal of the content by Meta.

Misinformation

Violations related to misinformation often involve the sharing of false content that can cause harm. This could include sharing misleading information about public health, such as promoting fake vaccine cures, circulating false claims about election processes, which can disrupt democratic systems, and creating and promoting conspiracy theories without any factual basis, potentially inciting fear or division. For example, a page that disseminates false information could result in page demotion or account suspension.

Harassment and bullying

Violations occur when users engage in repeated harmful behaviour against individuals or groups, which includes targeted campaigns to humiliate or degrade someone based on their appearance, gender, or personal decisions, posting private information like phone numbers or addresses with malicious intent (doxxing), and threats or degrading comments aimed at children or vulnerable individuals. For instance, a user repeatedly tags an individual in offensive posts, encouraging others to harass the victim. Meta intervenes by removing the content and potentially banning the user.

Impersonation and fake accounts

Impersonation violations involve creating accounts or pages that falsely represent individuals, brands, or organizations. This often includes creating fake profiles to scam others or spread disinformation, impersonating public figures or celebrities to defraud their followers, and using the identities of people to mislead others, such as during phishing attacks.

If an account pretending to be a charity organisation asks for donations through suspicious links, once reported, Meta will deactivate the account.

Child exploitation and sexual content

These violations are highly severe, involving content that sexualises minors or promotes harmful sexual behaviour. Violations include sharing images or videos of child exploitation or abuse, content depicting non-consensual sexual acts or sexual violence, and posts that solicit sexual favours or exploitative behaviour from others. A page sharing images depicting child abuse will result in immediate removal and reporting to authorities under Meta's strict child safety guidelines.

Copyright infringement:

Copyright infringement on Meta platforms (Facebook, Instagram) occurs when a user posts, shares, or distributes content protected by copyright without the permission of the copyright holder. This can include music, videos, images, graphics, written works, and streaming events, TV shows, or movies that are owned by another individual or entity without proper licensing or distribution rights.

Meta has strict policies to protect intellectual property rights, ensuring that creators maintain control over how their content is used. If content is posted without proper authorisation, Meta removes it in response to takedown requests or as a result of its automated content monitoring systems.

Violations of Meta's Community Standards can result in several penalties, ranging from content removal and temporary account suspension to permanent banning, depending on the severity and frequency of the violation.

How does Meta address copyright infringement?

Meta operates under the Digital Millennium Copyright Act (DMCA), which provides a legal framework for handling copyright disputes. Key aspects include takedown requests, where copyright holders can submit a takedown notice via Meta's Intellectual Property Reporting Form if their work is being used without permission. If Meta finds that content violates copyright laws, it removes the material and notifies the user, and account offenders that repeatedly infringe on copyright may face stricter consequences, including suspension or permanent banning.

Tools to protect intellectual property:

Rights Manager is a tool provided by Meta that allows copyright holders to upload and protect their content by automatically identifying and managing unauthorised uses across the platform. Reporting Mechanism, where users can also report content they believe infringes on their intellectual property directly through the platform.

What will happen after copyright infringement?

Meta immediately removes infringing content when it's reported or detected; repeated violations can lead to the suspension or permanent banning of accounts; and in extreme cases, copyright holders may pursue legal action against violators.

Contact Information:

For further assistance or to report a scam, please contact our incident response desk at:

Phone: +211920050106
+211921350435
Email: info@safetycomm.org
info@211check.org

Stay Alert. Stay Protected.
SafetyComm South Sudan and 211 Check
