## **Polaris**

## CODED LANGUAGE ON SOCIAL MEDIA

## HOW TO DETECT CSAM & HUMAN TRAFFICKING ON POPULAR SOCIAL MEDIA PLATFORMS

Identifying coded language related to domestic minor sex trafficking (DMST) online on social media platforms is difficult but possible. FIU analysts found that more so than any other type of DMST, DMST related to the production of Child Sexual Abuse Materials (CSAM) was most common. It should be noted here that human trafficking and CSAM are not synonymous.

While the consumption and dissemination of CSAM is abhorrent, only the production of CSAM may fall under the Federal definition of sex trafficking when "the person induced to perform such act has not attained 18 years of age."\* For example, social media accounts that advertise CSAM production (such as live or custom videos or photos) fall under the purview of human trafficking, but accounts that focus on consumption or dissemination of CSAM, which appear to be the majority of such accounts, do not. We provide the following executive summary of our findings with the caveat that the CSAM and DMST nexus from what we found is rather niche, and do not plan to continue to conduct research in this area given our organization's primary focus. Our research confirmed that many accounts suspected of involving DMST were difficult if not impossible to differentiate from accounts suspected of involving pimp-controlled adult sex trafficking.

## **Polaris**

CSAM continues to be prevalent on prominent social media sites, such as Twitter and Instagram. While these companies have taken some precautions against proliferating these materials, we found that users have altered or self-censored certain keywords and hashtags to circumvent platform-banned words to evade authorities.

Often, this censorship will involve censoring a vowel(s) in keywords (e.g., "p\*rn" instead of "porn" or "yung/yng" instead of "young" or replacing symbols with letters, such as "@ge" for "age").

We found that monitoring these types of keywords led to associations with additional and/or new keywords, and believe that iteratively monitoring known and new suspected keywords can lead to the exponential identification of new suspicious accounts potentially involved in CSAM.

We have created a non-exhaustive (as keywords are ever changing) list of these altered keywords and hashtags that users continue to use to signal involvement in CSAM, which financial actors can use to identify accounts involved in this activity.

Further, in our limited research we found that gift cards, CashApp, and PayPal were used as payment methods and common communication platforms outside of the immediate social media platform included Kik, Telegram, Discord, and Snapchat.