Pirates of Charity: Exploring Donation-based Abuses in Social Media Platforms

Bhupendra Acharya[†] Dario Lazzaro[$] Antonio Emanuele Cinà[$] Thorsten Holz[†]
[†]CISPA Helmholtz for Information Security, [$]Università di Genova
bhupendra.acharya@cispa.de dario.lazzaro@edu.unige.it antonio.cina@unige.it holz@cispa.de

Abstract

With the widespread use of social media, organizations, and individuals use these platforms to raise funds and support causes. Unfortunately, this has led to the rise of scammers in soliciting fraudulent donations. In this study, we conduct a large-scale analysis of donation-based scams on social media platforms. More specifically, we studied profile creation and scam operation fraudulent donation solicitation on X, Instagram, Facebook, YouTube, and Telegram. By collecting data from 151,966 accounts and their 3,053,333 posts related to donations between March 2024 and May 2024, we identified 832 scammers using various techniques to deceive users into making fraudulent donations. Analyzing the fraud communication channels such as phone number, email, and external URL linked, we show that these scamming accounts perform various fraudulent donation schemes, including classic abuse such as fake fundraising website setup, crowdsourcing fundraising, and asking users to communicate via email, phone, and pay via various payment methods. Through collaboration with industry partners PayPal and cryptocurrency abuse database Chainabuse, we further validated the scams and measured the financial losses on these platforms. Our study highlights significant weaknesses in social media platforms' ability to protect users from fraudulent donations. Additionally, we recommended social media platforms, and financial services for taking proactive steps to block these fraudulent activities. Our study provides a foundation for the security community and researchers to automate detecting and mitigating fraudulent donation solicitation on social media platforms.

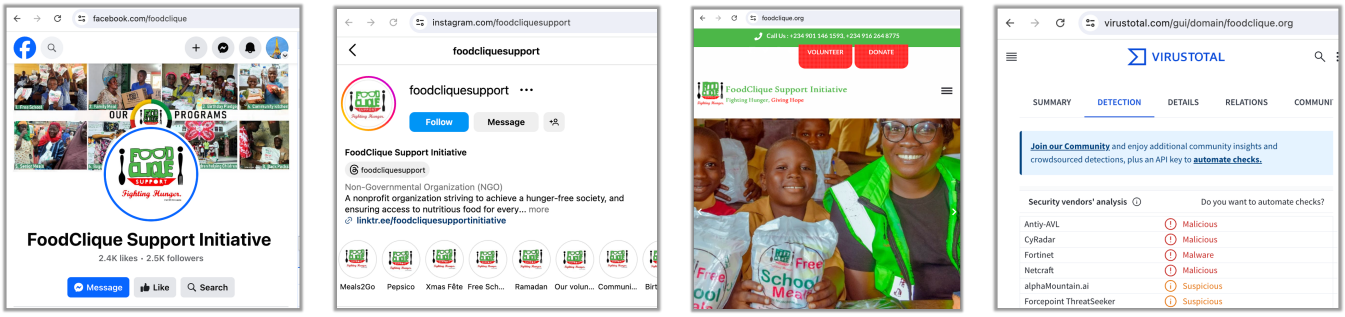
1 Introduction

Recently, there has been an increase in fraudsters using social engineering tactics to trick people into donating to fake charities or causes [1–3]. These tricks often include playing on sympathy and asking for a donation. Traditionally, fraudsters perform such attacks via the setup of fake donation websites [4], impersonation via phone calls [5], sending an e-mail or text asking to donate to a charity or cause [6], and sending a return letter envelope asking a cheque to send via mail [7]. As there has been a rise in social media users sharing, organizing, and participating in charity-related causes, this has simultaneously led to fraudsters shifting to conducting various donation scams on these platforms [8, 9]. Donation fraud, which is also commonly known as charity scam, is where scammers solicit money from individuals in the pretense of a charitable cause, disaster relief, or other seemingly legitimate reasons [1, 10]. These fraudulent activities can occur through various means, including fake websites, emails, social media posts, and crowdfunding platforms [2, 4, 11, 12]. The scammers deceive donors by pretending to represent real charities or by creating fictitious causes, often using emotional appeals to make urgent donations. Once the money is donated, it is typically diverted for the scammer's personal use, and the intended cause or individuals in need receive no benefit [2].

Over the years, social media users have steadily grown and are projected to reach 5 billion by 2025 [13]. Social media is popular among legitimate organizations and individuals to request donations for various causes [14]. It provides building networks and easy sharing for users and charitable organizations through posts, tags, and direct message communications [15]. Unfortunately, as social media adoption for donations has increased, fraudsters have also shifted towards social media-based donation scams. These scams include but are not limited to impersonating profiles of well-known organizations, individuals, or family members. Scammers often try reaching out by sending thank-you notes via direct messages, tagging posts for donations that users never made, or sending a friend or network requests to further establish a connection in the act of performing donation-based scams [16, 17].

According to the FTC, social media-based scams are on the rise, with more than 2.7 billion in losses from 2021 to 2023 [18]. Social media offers easy account creation compared to launching web domains, which often requires going through hosting websites and content. Various donation scams are increasing, with fraudsters posing as reputable organizations and soliciting contributions [3, 19, 20]. Scammers performing donation-based abuse in social media are ever rising [21–23], and with the rise of AI tools and content creation scammers are trending to abuse social media higher than before [24]. With the wide adoption of cryptocurrency globally, scammers are also shifting towards requesting donations via cryptocurrency [25–27] and using crypto drainers as part of the fraud. These crypto drainers trick victims into connecting through fake web wallet browsers, stealing their private key phrases, and ultimately draining the total funds from their wallets [28, 29]. In appendix Figure 1, we display an example of fraudulent donation soliciting on multiple platforms. Despite fraudulent donations being rampant on social media, there still lacks an end-to-end life cycle study of scammers' behavior, operation, and financial impact.

In this work, we address the research gap in donation-based abuses by conducting a study across five social media platforms. We assess profile creation, user engagement, and the external communication channels that scammers use to solicit contact and payments for fraudulent donation scams. Specifically, we conduct the first large-scale study of donation-based abuses on *X*, *Instagram*, *Telegram*, *YouTube*, and *Facebook*. Using donation-related search contexts, we collected data from 150K social media users and 3M posts. By analyzing the scammers' profile metadata and posts, including fraudulent emails, phone numbers, and URLs, we identified 832 scammers conducting fraudulent donation solicitations across these platforms. Additionally, our network analysis on these scamming accounts uncovers an additional 1K accounts linking to 11 platforms beyond their originating platforms. Furthermore, we provide an in-depth analysis of the scamming profiles' account creation, engagement posts, and techniques used to lure victims



(a) Facebook Donation Scam Profile Page **(b) Instagram Donation Scam Profile Page** **(c) Website Affiliated to Scamming Profile** **(d) VirusTotal Anti-Phishing Engine Fraud Analysis**

Figure 1: Examples of Scamming Donation Support Request: The first two images Figure 1(a), Figure 1(b) show the associated social media profile of the scamming donation on Facebook and Instagram social media platforms. The third image Figure 1(c) shows the associated external website asking for a donation to support and the last screenshot Figure 1(d) shows the risk engine evaluation from multiple anti-phishing engines (*Antiy-AVL, CyRadar, Fortinet, Netcraft, AlphaMountain.ai* and *Forcepoint ThreatSeeker*) indicating that the website is malicious or suspicious. The social media profiles can appear genuine, making it difficult to recognize the scam at first glance.

into fraudulent donations. Our findings show that social media platforms are not effectively blocking fraudulent accounts or protecting users against such abuses. Finally, we offer recommendations for proactive blocking and mitigation of these fraudulent activities for various platforms and payment processors.

Contributions. Our key contributions are as follows:

- **Fraudulent Donation Solicitation Measurement.** We conduct the first large-scale study of fraudsters soliciting donations across multiple social media platforms. Our approach uncovers scam accounts and their interconnected operations extending beyond their original platforms.

- **Fraudulent Payment Detection.** We identify fraudulent payment profiles and channels used by scammers to collect payments for fake donations. This enables tracking of financial losses and provides a blueprint for financial services to implement proactive solutions for detecting payment-related fraud.

Ethical Concerns and Data Disclosure. Our research did not involve interaction with any human subjects, including scammers. We collected public data from social media profiles using API queries. Additionally, we disclosed our findings to all five social media platforms: *X*, *Instagram*, *Facebook*, *YouTube*, and *Telegram*. For payment profiles linked to scamming accounts, we collaborated with *PayPal* and the cryptocurrency abuse database *Chainabuse*, both of which provided positive feedback and scam validation. We also shared email addresses, phone numbers, crowdfunding URLs, and survey forms associated with these scamming profiles with their respective service providers. *PayPal* confirmed that the flagged accounts were involved in various nefarious activities. *Chainabuse's* evaluation of cryptocurrency addresses revealed the scale of these attacks and associated financial losses. In summary, our work received several positive acknowledgments and validation of the abuses caused by fraudulent social media profiles soliciting donations. We provide our research code in a GitHub repository [30] to foster future research. However, data related to scammers will be only shared with the researcher upon request to prevent potential retribution attacks.

2 Related Work

In this paper, we perform a holistic study of scammers performing donation-based abuses across five social media platforms. To the best of our knowledge, we are the first to perform a large-scale analysis of donation-based abuses orchestrated by fraudsters on multiple platforms. Given the extensive research on scams and abuses over the past two decades, in this section, we focus on how our work diverges from previous studies and highlight the novelty of our approach in validating donation-based abuses.

Domains: Abuses, Scams, and Attacks Study. The use of web domains for distributing scams, and attacks remains a potent channel for abusers and has been widely researched over the last decade. These include studies such as traditional phishing attacks [31–36], Technical Support Scams [37, 38], and beyond such as Squatting-based attacks [39–44], and Malvertisement [45–47]. For instance, in PhishFarm [31], the author studied how malicious actors evade the anti-phishing engines in distributing various forms of scams and abuses in web domains. Agten et al. [42] studied squatting-based attacks that malicious actors perform via registering the squatting domains. With the rise of the adoption of digital currency over recent years, online frauds and attacks related to cryptocurrency scams are found ever rising, and tracking this fraud has caught the interest of security communities [48–51].

Social Media: Abuses, Scams, and Attacks Study. With the rise of abuses, scams, and attacks in social media platforms, social media has been a platform of interest to measure the prevalence of abuses among security communities and researchers. These studies explored various categories of social media scams including but not limited to Technical Support Scams [52], Comment Scams [53],

Cryptocurrency Abuses [54], Fake Profiles [55, 56] and Impersonation Attacks [57] revealing the widespread nature of these issues on social media. Abusers continuously develop new attacks, making detecting malicious profiles based on publicly available data has become increasingly challenging for the security community and researchers. For example: in HoneyTweet [52], Acharya et al. studied creating baiting tweets to lure scammers into an interaction with the posted tweets and performed an interaction with scammers to identify the modus operandi. The author also continued studying the variety of attacks that abusers perform as part of impersonating brands in the top 10K brands in multiple social media [57]. The most relevant work to us in areas of YouTube-based comments was studied by Li et al. [58], which analyzed scam campaigns and evasion techniques that scammers distributed as part of interacting comments on YouTube.

Donation Abuse Study. In areas of donation-based study, some of the prior work that are most relatable are from [59–62]. Whitty et al. [59] examined the psychological profiles of cyber scam victims and the types of scams associated with these profiles. Among these scams, one of the scams studied on charity scams involving fake profiles and organizations that deceive victims into donating to fraudulent causes. Korsell et al. [60] explored a taxonomy of fraud prevalent in 2020, highlighting the rise of charity and consumer scams. Similarly, Wood et al. [62] studied the various scams that were found emergent during COVID-19 and touched upon charity-based scams that were rampant during COVID-19. However, neither of these studies provided an in-depth analysis of how donation-based scams are propagated via social media platforms or the lifecycle of these scams as conducted by malicious actors.

Novelty. The prior work on social media has predominantly focused on other forms of attacks. Addressing this gap, our research performs an in-depth analysis of donation-based abuses on social media and their validation as scams. We leverage a straightforward methodology backed by LLMs and security risk engines well suited for identifying fraudulent profiles soliciting donations. The novelty of our work lies in identifying large-scale donation-based abuses across multiple platforms beyond the originating social media platforms and validating these scams through the association of fraudsters' payment profiles.

3 Evaluation Setup and Data Filtration

In this section, we detail our evaluation setup for identifying abusive social media profiles, particularly those soliciting fraudulent donations. We start by collecting data, including associated posts, from various social media platforms. This data is then filtered to focus on fraudulent donation solicitations, enabling a deeper analysis of scam operations. As shown in Figure 2, our measurement setup consists of three main components: ❶, which gathers data from various social media platforms using donation-related keywords; ❷, which filters the data to pinpoint profiles involved in donation scams; and ❸, which tracks the scammers' methods of operation. We provide details for each component as below.

3.1 Raw Dataset Aggregation

In order to aggregate the raw dataset, we perform two main tasks: (i) identifying relevant search keywords and (ii) conducting automated queries of the dataset across five social media platforms using these targeted search keywords. We provide further details below.

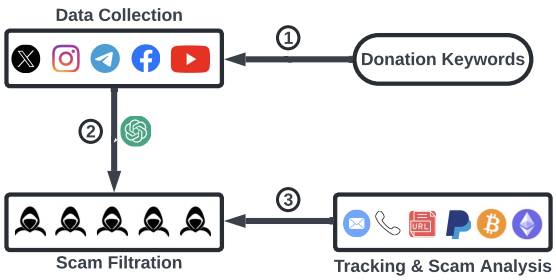


Figure 2: *Evaluation Setup Design*: An overview of our system, which consists of mainly three components: (i) *Data Collection* which performs automated donation-based keyword searches in five social media platforms, (ii) *Scam Filtration* which performs data filtration associated to donation soliciting fraudulent accounts, and (iii) *Tracking and Scam Analysis* which provides an evaluation of scammer's modes of operation and techniques.

Donation Keywords Identification. During our incubation phase, we manually reviewed online donation solicitations. We found 14 key terms frequently used in such solicitations, such as *givebetter*, *fund*, *help*, *act of kindness*, *support*, *charity*, *donate*, *donation*, *donor*, *awareness*, *giving*, *foundation*, *contribute*, and *helpsomeone*. These terms were linked with specific causes such as *cancer*, *earthquake*, *firefighters*, *police*, *veterans*, *animals*, *hunger*, *Ukraine*, *Christmas*, and *COVID-19*. Overall, we developed 78 keywords to search relevant posts and profiles across various social media platforms.

Data Collection. Utilizing API services[63–69], we gathered data across *X*, *Instagram*, *Facebook*, *Telegram*, and *YouTube* using the formulated keywords. We conducted three separate data searches for each social media platform from 2024-03-03 to 2024-05-15. In total, we collected 151,966 accounts and 3,053,333 posts from five social media platforms. Additionally, we retrieved profile metadata for each account, including name, description, links, profile image, timelines posts, and other publicly available information. A detailed breakdown of the raw data is presented in Table 1.

3.2 Fraudulent Donation Filtration

After collecting data from 151,966 accounts and 3,053,333 posts across five social media platforms, we conduct data curation. This process involves two primary steps: (i) pre-processing the raw data to confirm it pertains to donation-related contexts, and (ii) filtering candidates associated with donation-based abuses. The following outlines the various steps involved in our data curation techniques to ensure the accuracy of our findings.

3.2.1 Pre-Processing on Raw Data. In the pre-processing technique we perform filtrations by donation solicitation posts. During our manual analysis of the collected data, we found that API responses

Table 1: Overview of the raw dataset from five social media platforms. Our dataset reveals that *Telegram* has the highest number of accounts and posts compared to the others.

Social Media	Accounts	Posts
Instagram	1,604	136,082
Facebook	10,607	29,349
X	23,871	280,789
YouTube	30,482	54,314
Telegram	85,402	2,552,799
All	151,966	3,053,333

often contained irrelevant content. For example, searches using keywords like *donate cancer* yielded results that were not specifically about donations but included general cancer-related content or unrelated donation activities. To address this, we introduced a context check for each account and its associated posts to verify if the content was relevant to donation activities. Using the Large Language Model (GPT-4o) [70], we developed a prompt injection to identify whether a given post was relevant to the donation context (see Appendix A). This filtering process excluded accounts that were unrelated to the donation context: 25.56% (410/1,604) from *Instagram*, 79.84% (8,469/10,607) from *Facebook*, 20.77% (4,959/23,871) from *X*, 80.93% (24,670/30,482) from *YouTube*, and 89.12% (76,111/85,402) from *Telegram* were filtered. Across all five social media platforms, this filtering removed 75.42% (114,619/151,966) of accounts and 82.45% (2,517,489/3,053,333) posts associated with these accounts from our raw dataset. We then applied security risk engine-based flagged association to the remaining 24.57% (37,347/151,966) accounts and their 17.54% (535,844/3,053,333) posts related to the donation context to identify candidate scam accounts.

3.2.2 Data Filtration and Labelling. To label an account as a donation solicitation scam, we apply two criteria: (i) the account solicits donations through publicly engaged posts, and (ii) the account's communication channels or profile metadata include elements flagged by security risk engines. If both conditions are met, the account is labeled as a candidate for donation solicitation scam. For example, if a social media profile solicits donations and includes a fraudulent email, phone number, or links to websites flagged by Anti-Phishing Engines as phishing URLs or malicious emails, we categorize it as a donation solicitation fraudster. Further details on the filtering and data labeling techniques are provided below.

Phishing URLs. We observed that social media profiles often include external websites or URLs in their bio sections. For each profile, we analyze the metadata to check for the presence of any URLs or domains. Using the *VirusTotal* API [71], we evaluate whether these URLs are flagged as phishing or scam sites. To ensure accuracy, we only consider URLs or domains as potential candidates if they are flagged by at least two security risk engines from *VirusTotal*. Accounts or posts containing URLs flagged by *VirusTotal* are marked for further scam donation abuse analysis.

In total, we identified 118,735 URLs within the profile metadata, and 0.95% (1,128/118,735) of these distinct URLs were flagged by at least one of the *VirusTotal* security risk engines, spanning 2,345 social media accounts. Of the 1,128 flagged URLs/domains, only 22.34% (252/1,128) were flagged by two or more security risk engines. A manual review of 5% of the URLs, both single-flagged and multi-flagged, revealed that single-flagged URLs/domains were often false positives or unknown, while those flagged by two or more engines were found to be reliable. To mitigate potential false positives, we labeled accounts containing 0.21% (252/118,735) of URLs/domains as candidate accounts linked to 369 social media profiles that were flagged by multiple security risk engines from *VirusTotal*.

Abusing Email Addresses and Phone Numbers. We observe that social media profiles often include communication methods such as email addresses and phone numbers in their bio-data to facilitate user contact. To assess the reliability of these communication methods, we used third-party API services to check the fraud score of the provided email addresses [72] and phone numbers [73]. Social media profiles with communication methods having a fraud score greater than 85% were marked as candidates for further analysis. We set an 85% threshold based on the providers' high-risk validation, which indicates a strong association with fraud or high-risk activity for the given account. Out of 7,752 email addresses found in our pre-processed data, 2.90% (225/7,752) distinct email addresses were flagged with high-risk / fraud emails associated with 257 social media accounts. Similarly, out of 9,791 phone numbers found in our pre-processed data, we identified 1.37% (135/9,791) fraud phone numbers associated with 201 social media accounts.

In a nutshell, starting with 151,966 accounts and 3,053,333 posts from five social media platforms, we applied two filtration techniques: (i) Initially removing non-donation-based contexts, and (ii) Further curating the data based on fraud risk engine-flagged URLs/domains, phone numbers, and emails. As a result, our dataset for donation-based scams includes 832 social media profiles. This means we filtered out 99.45% (151,134/151,966) of the accounts from raw dataset accounts. We acknowledge that our conservative filtering approach may have excluded some donation scam accounts. However, as pioneers in the large-scale study of fraudulent donation scams, our goal was to build a solid foundation using known seed data to reduce potential false positives. Additionally, in section 11, we explore the data evaluation and the efficacy of scam filtration of our approach.

3.3 Tracking and Scam Analysis

The third component, tracking and scam analysis, focuses on evaluating data from scammers' profile metadata and engagement posts. We analyze profile metadata to investigate the scammers' associations with flagged email addresses, URLs, and phone numbers identified by fraud detection engines. Additionally, we examine engagement posts to understand scammers' interactions and operational methods to show how scammers solicit donations via financial payment methods such as *PayPal*, cryptocurrency addresses, survey forms, and crowdfunding services. By analyzing data from these sources, we provide details on scam operations and the connections between scam accounts across multiple platforms beyond originating social media platforms.

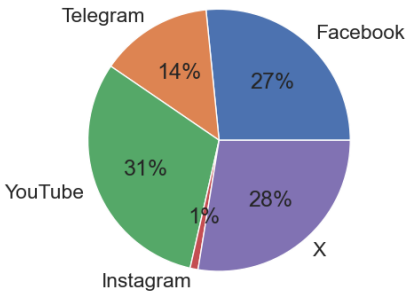


Figure 3: Distribution of security risk engines flagged communication channels (email, phone number, and URLs) across social media platforms. In this pie chart, we show the total number of scamming channels that were flagged by security risk engines identified across five social media platforms, with 31% of the total communication channels accounting from the *YouTube* platform.

For the rest of the section organization, we provide – an overview of donation abuse in section 4; profile content and association in section 5; fraudulent donation solicitations topologies in section 6; evaluation of scammer's profile picture in section 7; sentiment analysis of interacted comments in section 8; scammer operations and network analysis in section 9; and tracking of scamming payment profiles in section 10. Additionally, we provide recommendations for mitigating and proactively blocking these fraudulent accounts in section 12.

4 Scam Donation Abuse Overview

In this section, we provide an overview of fraudulent communication channels collected from scammers' profile metadata and engagement posts. In Table 2, we summarize these findings by social media platform. The first column lists the five social media platforms we studied. The second, third, and fourth columns show the number of fraudulent channels associated with the scamming accounts. The fifth and sixth columns provide the distinct and total posts identified in the context of donation scams, and the seventh column shows the overall number of scammers soliciting donations.

In total, we identified 225 fraudulent emails, 136 fraudulent phone numbers, and 252 malicious URLs shared by 832 scammers across 17,730 posts and profile metadata. Among these fraudulent communication channels, scammers showed a strong preference for URLs, which accounted for 41.10% (252/613), often directing victims to external websites for donations. The remaining channels included emails at 36.74% (225/613) and phone numbers at 22.21% (136/613). In Figure 3, we illustrate scamming channels by each social media profile, and below, we highlight key findings for each platform studied.

Instagram. In our study, 6.73% (56/832) of scammers operated on *Instagram*, the lowest count among the platforms analyzed. These scammers preferred using malicious URLs for donation fraud over emails or phone numbers. Among the 56 scamming accounts, we found no fraudulent emails, one fraudulent phone number, and 5 malicious URLs, which appeared in 25.97% (4,606/17,730) of posts and profile metadata. Notably, these scammers frequently duplicated posts to solicit donations; of the 4,606 posts reviewed, 78.57% (3,619/4,606) were duplicates.

Facebook. Among the five social media platforms, although *Facebook* had the second-lowest number of scammers at 19.71% (164/832) and the fewest posts at 1.82% (323/17,730), it accounted for the highest percentage of fraudulent emails—57.19% (147/257) of all identified fraud communication channels. This suggests that scammers on *Facebook* were more inclined to engage in donation-based fraud through emails rather than using fraudulent phone numbers or malicious URLs.

Telegram. In our study, *Telegram* had the second-highest post count at 34.26% (6,075/17,730) and accounted for 22.59% (188/832) of the scammers. Among the 85 distinct fraudulent communication channels linked to these 188 scamming users, phone calls were the preferred method, making up 74.11% (63/85). Since *Telegram* is widely used for text messaging and phone calls, scammers on this platform were most likely to connect with victims through phone calls or direct messages.

YouTube. On *YouTube*, 24.03% (200/832) of scammers were identified, the second-highest after *X*. Among the 190 fraudulent communication channels used by these 200 scammers, external URLs were the most common, accounting for 31.16% (115/369). Emails followed as the second most used method at 20.88% (47/225), with phone calls close behind at 20.58% (28/136).

X. Overall, our study found that the *X* platform is the most favored among scammers, comprising 26.92% (224/832) of all scammers. Among the 169 fraudulent communication channels identified on *X*, 67.45% (114/169) were malicious URLs, making them the most common method for donation abuse. Similarly, 36.80% (6,526/17,730) of the scamming posts featured a significant proportion of malicious URL sharing at 45.23% (114/252). The findings indicate that fraudulent profiles on *X* prefer using malicious URLs over emails and phone numbers to solicit fake donations.

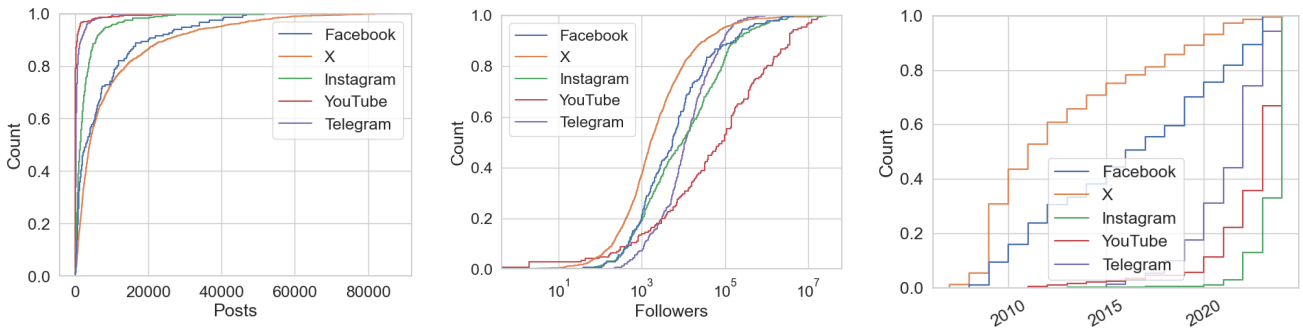
Key Takeaways. Through the study of abusive communication channels, we identify that scammers use social media platforms as originating sources, and direct victims to use external channels such as fraud email, phone calls, and URLs to further contact. As URLs provide easy fraud mechanics compared to email and phone calls, scammers prefer URLs as the highest compared to others asking victims to donate via external sites.

5 Profile Content and Association

In this section, we dive deep into scammers' techniques to create profiles that attract potential victims on social media platforms. We conduct a thorough analysis of six key aspects: post engagement, follower count, account age, location settings, categorical representation, and account monetization. In Figure 4, we present a CDF graph showing scammers' engagement through posts, follower count, and account creation dates, and below we provide further details on profile content and associations.

Table 2: Summary of scammers' posts and communication channels. This table shows our findings on donation-based abuses identified by analyzing profile metadata and engagement posts from five social media platforms. For each communication channel—email, phone, and URLs—we perform queries to determine if security risk engines flag the communications.

Platforms	Fraud Email/Accts.	Fraud Phone/Accts.	Malicious URL/Accts.	Distinct Posts	Total Posts	Scammers
Instagram	0	1/12	5/44	987	4,606	56
Facebook	147/148	12/14	4/4	322	323	164
Telegram	8/8	63/84	14/86	6,049	6,075	188
YouTube	47/78	28/58	115/70	180	200	200
X	23/23	32/33	114/165	6,520	6,526	224
Total (Distinct)	225/257	136/201	252/369	14,058	17,730	832



(a) Scammer Interactions. **(b) Following Count.** **(c) Profile Creation Date.**

Figure 4: CDF Engagement and age of scammer profile from each of the social media platforms – Figure 4(a) shows the engagement of scammer via posts, Figure 4(b) shows the *following* count of scammers and Figure 4(c) shows the age of scammers based on profile creation date from each of the social media platforms that we studied.

Description/Bio. Scammers engaged in fraudulent donations were found to use various tactics in their profile descriptions. These descriptions provide a brief message to visitors. We found that 95.31% (793/832) of scammers contained profile descriptions that included messages related to emotional manipulation, credibility, authentication, details about the donation campaign, or appeals to generosity. The remaining 4.68% (39/832) of scammers were found to lack any description or bio information.

Posts Engagement. Out of 17,730 posts collected from five social media platforms, the overall median post interaction across all platforms was 709. The median post interactions for each platform were: *X* (4,775), *Instagram* (1,307), *YouTube* (2), *Telegram* (147), and *Facebook* (2,960). Our results indicate that scammers are more likely to engage on *X* compared to other platforms, whereas *YouTube* is less favored for engagement through posts.

Followers Engagement. The median follower count across the five social media platforms was 3,345. For each platform, the median follower counts were - *X* (1,621), *Instagram* (9,361), *YouTube* (82,500), *Telegram* (10,809), and *Facebook* (5,449). Our result indicates that users are more inclined to follow scammers in video-based donation contexts compared to post-based ones. Since videos are generally more engaging than posts, scammers may find it easier to attract and retain followers through video content.

Account Age. Our analysis of fraudulent social media profile ages reveals that scammers often use either harvested or aged profiles. The median creation date for all social media profiles was 2018. Specifically, the median ages for each platform were: *X* (2011), *Instagram* (2024), *YouTube* (2023), *Telegram* (2022), and *Facebook* (2016). This indicates that scammers are more likely to utilize older accounts on *X* while utilizing newer accounts on Instagram.

Location. We identified 50.12% (417/832) of scammers with 210 distinct geo-location sets as part of their profile information. The top three countries represented were *Russia* (62), the *USA* (37), and *India* (29). It is important to note that geo-location is an optional field and does not necessarily reflect the scammers' actual locations, as it is often populated with unrelated names. For example, the location name *global*, although not a real location, had the highest count, with a total of 77.

Categorical Representation. We identified that 49.87% (415/832) of scamming accounts featured 115 distinct categories or affiliations in their profiles. Among these, the top three categories included: *Non-Profit Organization* (62), *Charity Organization* (22), and *Non-Governmental Organization (NGO)* 17. The remaining 50.12% (417/832) were found to display missing categorical representation.

Account Monetization. We found that 39.66% (330/832) of scamming accounts across four social media platforms (*Facebook* (22/164), *Instagram* (30/56), *X* (218/224), and *YouTube*) (60/200), opted for

business or advertisement features. This enables these accounts to monetize their presence and allows the platforms to display advertisements. Accounts with higher engagement levels generally gain more from opting into these business features. Notably, 97.32% of scamming accounts on the *X* platform had the highest participation in business or advertisement features.

Key Takeaways. Scammers were found to use older social media accounts to launch donation abuse campaigns. We suspect these are rather harvested accounts. Scammer's geolocation data shows diverse representations of top countries including Russia and the USA, though these locations are often misleadingly named. Moreover, scammers often masquerade under popular affiliations and opt-in for business/advertisement features, allowing for monetization through advertisements.

6 Fraud Topologies: Anatomy of Fraudulent Donation Solicitations

In this section, we provide the categories of fraudulent donation solicitations that scammers perform through posts. We provide a technical overview and the findings of the scam clusters below.

Technical Overview. We clustered donation solicitation posts from 832 scamming profiles excluding non-English content. In total, we analyzed 17,706 posts across five platforms: *X* (6,526), *Instagram* (4,583), *Telegram* (6,075), *Facebook* (322), and *YouTube* (200). For language identification and filtering, we use the CLD2 library [74]. We then vectorized the posts using the *all-mpnet-base-v2* sentence transformer model [75]. Subsequently, we processed the posts using the BERTopic library [76] to remove redundant information, such as stop words. We combined UMAP [77] and HDBSCAN [78] for clustering, followed by the KeyBERT [79] model to refine topic representations within each cluster.

In the hyperparameterization process for UMAP, default values from the BERTopic library [76] were employed. Specifically, we configured UMAP with `n_neighbors=15`, `min_dist=0.0`, `n_components=5`, and cosine similarity. We then set the `random_state` variable to a fixed value of 42 to preserve the reproducibility of our code.

For HDBSCAN, we chose `min_cluster_size=10` and used the Euclidean metric for clustering. To refine the clustering outcome, we adjusted `min_samples=50` to reduce the resulting number of clusters. Additionally, the default BERTopic method for outlier reduction (`reduce_outliers`) was applied to minimize the presence of outlier samples in the clustering results. Finally, we employed a standard evaluation metric, i.e., silhouette score [80], and visual inspection of resulting clusters to assess the quality and validity of the clustering outcomes.

Clustering Results. We conducted a manual qualitative analysis of prominent scam categories identified in our findings. Out of the 62 clusters identified through our clustering pipeline, we present below an analysis of the top 10 clusters based on engagement through posts where scammers solicit fraudulent donations.

- **Urgent Support.** We observe that scammers frequently target specific donation days or weeks to create a sense of urgency, often setting rapidly approaching deadlines. A common tactic involves urging social media users to complete survey forms or to visit an external website before the donation period ends. We identified 185 scammers asking for urgent support fraudulent donations through 951 posts, which comprised the highest numbers of scammers and post-interactions in our study.
- **Animal Rescue.** In the context of animal rescue abuse, scammers target individuals by posing as representatives of legitimate animal rescue organizations to establish credibility. These fraudulent posts solicit donations under the guise of supporting animal welfare causes, asking for contributions to help save and care for animals in need. In this cluster, we identified 125 scammers asking for fraudulent animal rescue donations via 679 posts.
- **Disaster Relief.** We observe scammers often exploit the impact of disaster relief to solicit fraudulent donations. In this category, scammers act as legitimate organizations or affiliations preying on those looking to support natural disaster victims. We identified 87 scammers asking for disaster relief fraudulent donations through 426 posts.
- **Event and Activities Support.** Scammers in this category exploit popular events to solicit fraudulent donations, leveraging the excitement and urgency to support the occasions. The scammer was often found to craft persuasive messages appealing to participants' emotions and sense of community, urging them to contribute financially. In this cluster, we identified 80 scammers asking for fraudulent donations via 427 posts.
- **Crypto Scams.** In the context of crypto donation abuse, we identify that scammers exploit the growing popularity and perceived anonymity of cryptocurrency to solicit fraudulent donations. They take advantage of the novelty and complexity of cryptocurrency, making it appealing for users to either participate in charity-related philanthropic support or take part in free crypto token giveaways. In this cluster, we identified 57 scammers asking for fraudulent crypto donations via 111 posts.
- **Holiday/Seasonal Spirit.** Scammers in this category exploit the holiday or seasonal spirit of generosity to make fraudulent donation requests. These scams are often focused on children and families in need. In this cluster, we identified 54 scammers soliciting fraudulent donations through 382 posts.
- **Education/Research Support.** In this category, we observe scammers exploit the education sector by targeting individuals with fraudulent donation requests related to scholarships, educational research, and student support. These scams often pose as associations to institutions or charitable initiatives, appealing to the goodwill of alumni, faculty, and the general public. We identified 49 scammers soliciting fraudulent donations through 92 posts in this category.
- **Ticketing and Offer Exchange.** We observe that scammers in this category claim to need tickets or offer ticket exchanges as part of fraudulent ticket donations. Scammers perform potential disguises as potential donors to fraudulent

websites or request personal information under the guise of facilitating a ticket donation. By creating a sense of urgency and community solidarity, they deceive well-meaning fans into buying a sold-out ticket or providing financial support for a particular event through ticket purchases. In this cluster, we identified 46 scammers asking for fraudulent ticket donations via 81 posts.

- **Narcissistic Abuse Support.** In this category of fraudulent donation solicitations, scammers target individuals by asking for support for abused groups, particularly those affected by narcissistic abuse. Their tactics include raising awareness and soliciting donations for victims of war, domestic violence, and psychological abuse. In this cluster, we identified 40 scammers asking for fraudulent donations via 63 posts.
- **Medical.** In medical-related fraudulent donation requests, scammers are found to solicit funds for various medical causes, such as covering the medical expenses of a critically ill patient, supporting medical research, or providing medical care for disadvantaged groups. They often impersonate medical institutions to add legitimacy to their appeals. In this cluster, we identified 36 scammers asking for fraudulent medical-related donations via 265 posts.

Key Takeaways. Our analysis of post-clustering uncovered several scam categories of fraudulent donations performed by social media profiles. These include urgent appeals with specific deadlines, schemes tied to events, holiday-themed solicitations for families and children, and deceptive campaigns masquerading as education and research support. Furthermore, scammers exploit disaster relief efforts, victims of abuse, animal rescue, and medical issues, presenting themselves as legitimate fundraisers while seeking fraudulent donations.

7 Evaluation of Scammer Profile Picture

In this section, we provide an evaluation of the scammer's choice of profile picture while soliciting donations across multiple social media platforms. We provide a technical overview and the findings of the profile picture evaluation below.

Technical Overview. Using unsupervised clustering to identify patterns and relationships among these images, we examine the profile pictures of scammer accounts. Following the methodology outlined in [52], we collected profile pictures and employed the pre-trained visual model CLIP [81] for feature extraction. For each profile picture, we extracted the CLIP token embeddings and rescaled the images to a resolution of 224×224 pixels to match the input size used during the model's training [81]. These embeddings were then visualized using Uniform Manifold Approximation and Projection (UMAP) [77]. To identify clusters and eliminate anomalies, we applied standard clustering algorithms: HDBSCAN [78] and single-linkage hierarchical clustering [82]. Below we provide additional detailed information on the chosen hyperparameters and clustering validation, and the results of our findings.

Table 3: Clustering analysis of scammers' profile pictures and their distribution across the five social media platforms. Our result reveals that scammers in the Association-Logos category were found to be the highest and utilize association logos to solicit donations.

Cluster Label	Count	Facebook	X	Telegram	Instagram	Youtube
Associations Logos	240 (29.13%)	79	102	34	24	1
Male/Female	133 (16.14%)	14	48	42	12	17
Video Clips	110 (13.35%)	0	0	3	1	104
Games & Cartoon	103 (12.50%)	12	33	49	1	7
Politics/War	97(11.77%)	0	0	33	1	63
Pets	85(10.31%)	48	17	0	16	4
Low-Resolution	37 (4.49%)	11	16	9	1	0
Crypto Coins	19(2.31%)	0	3	16	0	0
Total	824	164	219	186	56	196

Clustering Hyperparameters Selection. During hyperparameters selection, we employ standard evaluation metrics, i.e., silhouette score [80] and Calinski-Harabasz score [83], and visual inspection of resulting clusters to assess the quality and validity of the clustering outcomes. For both UMAP and DBSCAN, we systematically tuned their hyperparameters to optimize clustering pipeline performance and obtain meaningful and reliable results. To this end, we considered a wide range of hyperparameter configurations. Specifically, for UMAP, we let the `n_neighbors` hyperparameter vary in the intervals $[3, 100]$ and set the `n_components` equals to 2 to visualize the clusters. Regarding DBSCAN, we let the `min_cluster_size` and `min_dist` vary in the intervals $[5, 100]$ and $[1e-02, 1]$ respectively. The resulting investigation involved $2,500$ configurations of these hyperparameters, identifying the configuration `n_neighbors`=15, `n_components`=2, `min_dist`=0.1, and `min_cluster_size`=20 as the most reliable, according to their silhouette and Calinski-Harabasz scores, for our clustering pipeline.

Clustering Results. We present the results of our clustering analysis on 824[1] scammer profile images in Table 3. From the analyzed dataset, we identified seven common categories of profile pictures used by scammers: *Association Logos*, *Male/Female*, *Video Clips*, *Games & Cartoon*, *Politics/War*, *Pets*, *Low-Resolutions*, and *Crypto Coins*. Our results show that 29% of scammers use *Association Logos* as their profile pictures, often featuring logos from various groups such as pacifist organizations, religious institutions, private companies, or even the *Ukraine* flag. About 16% of scammers use *Male* or *Female* profile pictures, while 13% fall into the *Video Clips* category, using video snapshots as their profile images. The *Games & Cartoon* category, comprising 12% of scammers, includes images of video game characters, anime protagonists, and memes. Additionally, 11% of scammers employ *Political War* images, such as screenshots from political news, military actions, or propaganda. The *Pets* category (10%) features images of animals, mostly cats and dogs, as well as pet-related activities. Scammers using low-quality images belong to the *Low-Resolution* cluster (4%), where the content is difficult to discern. Finally, 2% of scammers fall into the *Crypto Coins* cluster, which includes images of cryptocurrencies, and wallet logos.

[1]We excluded 8 images due to unsupported formats (e.g., non-JPEG or non-PNG)

Our analysis of scammers' profile images revealed that they often aim to emotionally manipulate users by featuring images of pets, war, educational organizations, or religious themes. Additionally, in Appendix, Figure 6-8, we show a subset of 50 scammer profile pictures from *Association-Logos*, *Male*, *Female*, *Games-Cartoon*, *Politics-War*, *Pet Associations*, and *Pets* clusters. Notably, the content within the clusters we identified is cohesive and coherent with our assigned label. Complementary, in Figure 9, we illustrate samples coming from the *Miscellaneous* cluster, which contains a mixture of pictures that have been considered anomalous by our clustering algorithms.

Key Takeaways. Through profile image analysis, we identify patterns and tactics used by scammers to create a deceptive online presence. Our analysis revealed that scammers predominantly use association logos, male and female images, political war, and game/cartoon characters to appear credible. Such insights are valuable for developing targeted measures to detect and counteract fraudulent activities, improving online security across social media platforms.

8 Sentiment Analysis of Public Comments

In this section, we conduct sentiment analysis between users and scammers. We focused specifically on YouTube due to its unique video-based interaction format. Users often engage with videos as directed by the content, which differs from textual posts found on posts-based interacting platforms (*X*, *Instagram*, *Facebook*, and *Telegram*). We collected 3,676 distinct comments from 364 scamming YouTube channels.

Technical Overview. For sentiment analysis, we utilized the Llama3-8B model based on its popularity as the start of an art open-source model on benchmark sentiment analysis. Our comment categorization was based on predefined sentiments: *Gratitude*, *Action*, *Anger*, *Abuse*, and *Neutral*. We provide the prompt detail to these five sentiments in Figure 5.

Sentiments Results. We provide detailed results of sentiment analysis of post engagement between users and scammers during the lifecycle of fraudulent donation solicitations as below.

Gratitude. In the *Gratitude* category, we measured comments expressing gratitude, relief, or thankfulness. We found that 53.73% of the comments reflected gratitude. We observe that scammers frequently try to thank those who have already donated and solicit others to make additional fraudulent donations. An example of a scammer's gratitude is shown below.

Thank you, every single donation matters, even if you can't donate.

God bless everybody involved in the rescue and care of this beautiful dog family.

7 hours and already $50,000 donated... Thank you for improving the lives of so many others.

```
You are a classifier. Given a Comment, classify it into one of the
↪ following categories:

- Gratitude : A comment expressing gratitude, relief, or similar
↪ emotions.
- Action : A comment that includes awareness, a report, an urgent
↪ action, or similar prompts.
- Abuse : A comment indicating that scammers are engaging in
↪ hateful, abusive, fearful, or concerning activities.
- Anger : A comment showing that the user is frustrated or angry
↪ because they believe YouTube is not taking serious steps to
↪ block scam accounts.

Please provide the category name and a brief explanation for your
↪ classification in the following format:
Category: "..."
Explanation: "..."

Examples:

Comment : "Thank you so much for addressing this issue! I was
↪ really worried."
Category : "Gratitude"
Explanation : "The comment expresses gratitude and relief for
↪ addressing the issue."

Comment : "Everyone needs to report these scammers immediately!"
Category : "Action"
Explanation : "The comment is a call to action, urging others to
↪ report scammers
```

Figure 5: System prompt for Llama-3. We instruct-tune Llama-3-8B to classify sentiment in Youtube users comments with a system prompt describing the task and two examples.

Action. In the *Action* category, we measured comments that include awareness, report, urgent action, or time-based responses. We found that 17.79% of the comments interacted with scamming videos displayed action. We provide examples of action below.

Donate please, another 7.4 earthquake struck Nepal just now. Quality of life and hospice support is imperative. Now that you learned how to make a donation button in PLS DONATE.

Most large charities are scams with a fraction of donated money ever reaching those it was gifted for, give to local charities that actually do good work.

Anger. In *Anger* category, we measured comment that shows that the user is frustrated or angry because YouTube does not take serious steps in blocking the scamming accounts or scammers. We found that 16.43% of the dataset typically showed frustration with YouTube's handling of scam accounts.

They need to be closed down and thrown in jail for fraud.

The scam part angers me.

*Contact us about paying them for their scam a** service.*

Hate. In *Hate* category, we measure engagement in hateful or abusive behavior on interaction. We identified 11.62% of the highlighted

engagement comprised of hateful or abusive behavior. Examples of such hateful comments from scammers are shown below.

You hate charity because youŕe a cringe Socialist.

*You should get out there on the streets and do the fuc**ng work.*

You're a lying imposter you deserve what misfortune that comes your way.

Neutral. In *Neutral*, we measure interaction that is not necessarily related to donation-based context or posts. We suspect these neutral comments are rather scripted to gain followers. We found the neutral context as the lowest category comprising 0.43% of our overall dataset. An example of a neutral comment are shown below.

Fun fact: snakes actually use their tongues to catch scents!

Line from Seinfeld: G̈eorge likes his Kung Pao SPICY¨.

*If you are impressed with this video, please support us on Patreon - https://www.patreon.com/Le**cs. It will be a great help for us.*

Key Takeaways. Our analysis of scammer and user interaction sentiments revealed several key insights. In the *Action* category, comments reflected urgent responses or awareness, with some users advising against taking action due to mistrust of large charities. The *Anger* category showed that comments expressed frustration with YouTube's failure to block scam accounts. In the *Hate* category, interactions involved hateful or abusive behavior, both from scammers and users. Lastly, the *Neutral* category included unrelated, scripted comments and motives to gain followers. This indicates that comment-based interactions are lucrative channels of operations for scammers, offering interactive video-based solicitations for donations.

9 Scammer Network Analysis

In this section, we explore how scammers operate across multiple social media platforms, focusing scam cycle and modus operandi. We detail the fraud lifecycle, illustrating how scammers redirect users from one platform to another through tactics such as crowdsourcing, and external links, and share scam channels across multiple profiles. We provide further details below.

9.1 Operation Beyond Originating Platform

In this section, we specifically focus our analysis on scammers operating beyond the originating platforms and interlinking accounts among multiple platforms. Our analysis primarily covers (i) external platforms that scammers link to their profiles, (ii) donation solicitations via crowdfunding services, and (iii) survey forms. Below, we provide detailed information on each category.

External Communication Channels. Through profile metadata analysis, we found that scammers frequently include details of external platforms in their bio descriptions, linking them to the originating social media platform. We identified two types of external bio links on scamming profiles. The first type links to external websites such as Linktree URLs, which aggregate multiple platforms and related links to the scammer's account. For example, a bio profile linking to *www.linktree.com/scam_account_1* was often found to contain various social media accounts associated with scam accounts, such as *www.facebook.com/scam_account_f*, and *www.twitter.com/scam_account_t*. The main purpose of these accounts is to provide visitors with a choice of platforms for contact. The second type involves direct links to a preferred platform, such as an *X* profile containing links to *Instagram* or *Telegram* as part of the external contact details.

Table 4: Overview of the external platforms linked to the scam accounts. In this table, we show scammers interlinking various platforms as part of a scam operation.

Social Media	External Linked Accounts
YouTube	482
Instagram	166
Facebook	122
Twitter	83
Amazon	44
LinkedIn	36
TikTok	30
Telegram	25
Etsy	9
Signal	2
WhatsApp	2
All (Distinct)	1,001

We observed that 37.5% (312/832) of scamming accounts included external links in their profiles, with 127 of these accounts linking multiple bio profiles (ex. *Linktree*) to external websites. For accounts with multiple external bio links, we automated the Selenium Python script to gather the associated platforms interlinked with the originating account. In Table 4, we present data showing 832 scamming accounts interlinked with 11 different platforms across both categories. Overall, we identified 1,001 distinct external platform accounts linked beyond the study accounts. Among these platforms, the top five most commonly interlinked accounts were related to *YouTube* (48.15%), *Instagram* (16.58%), *Facebook* (12.18%), *Twitter* (8.29%), and *Amazon* (4.39%). To gain further insights, we conducted a manual analysis by randomly selecting 100 accounts and visiting each link through a browser. We identified four distinct scam operation techniques: (i) platforms such as YouTube were used for video-based donation requests, (ii) messaging platforms such as *Signal*, *Telegram*, and *WhatsApp* were used for direct communication, (iii) social media platforms like *Twitter*, *Instagram*, and *Facebook* were primarily utilized for post engagement, and (iv) consumer-oriented platforms such as *Amazon* and *Etsy* were exploited by scammers to solicit support through purchases from wishlists or gifts. Thus, starting with 832 scamming accounts from five social media platforms, this technique yielded an additional 1,001 external accounts linked to 11 platforms (9 social media platforms and 2 online e-commerce platforms).

Table 5: Overview of the crowdfunding services. Our results show that scammers often redirect users from the original social media platforms to seven crowdfunding services.

Crowdfunding Services	Scam Accounts	Fund Links
Patreon	45	40
Givebutter	24	6
Donorbox	6	19
Kickstarter	4	8
Indiegogo	2	2
Fundrazer	1	1
Rallyup	1	1
All (Distinct)	83	77

Crowdfunding Services. We found that scammers exploit crowdfunding services for donation solicitations. We analyzed the presence of popular crowdfunding service URLs in posts engaged by scammers. As shown in Table 5, we identified 9.97% (83/832) of scammers soliciting donations via 77 URLs from seven different crowdfunding services. The top three platforms used were *Patreon* (51.94%), *Donorbox* (24.67%), and *Kickstarter* (10.38%). We conducted a manual review of these 77 URLs by visiting each link in a browser. Out of 77 distinct URLs, 9 links were either inactive or deleted. Among the active URLs, 23/68 had already closed their fundraising campaigns, with amounts raised ranging from $25 to $58,180. The remaining 45/68 crowdfunding URLs were found to be actively collecting donations, using three main solicitation methods: (i) minimal payments to join a group as a form of support for the cause (e.g., *Patreon* memberships starting at $1.70 per month plus tax), (ii) recurring donations such as monthly or annual contributions ($5, $25, or higher), and (iii) one-time payments for support (ranging from $5 to several hundred dollars). Our analysis from the last week of September 2024 identified 3,696 contributors who donated over $252,620 through 37 active fundraising links. This amount does not include contributors who may have made or are still making donations via membership subscriptions. We suspect scammers are repeatedly defrauding victims through ongoing solicitations observed in our dataset.

9.2 Campaign Detection

We analyzed shared communication channels, specifically URLs, emails, and phone numbers, among scam accounts to determine whether these channels interlink scam accounts as part of their communication with potential victims. To do this, we aggregated data from abuse candidate scam accounts across all five social media platforms. If a minimum of two scam accounts share a single communication channel, we refer to the given group as a scam campaign shared by the scam accounts.

We grouped the scam accounts based on individual types of communication channels, such as emails, URLs, or phone numbers. In Table 6, we summarize the scam clusters, including the minimum, maximum, and median counts of scam accounts per cluster. Overall, 42.66% of scam accounts were found to be part of scam campaigns. Among these, URL clusters were the most prevalent, with 41 distinct clusters comprising 231 accounts, while email clusters were the least common, with 12 distinct clusters involving 44 scam accounts.

Table 6: Overview of scammers sharing the communication channels. The table provides a breakdown of clusters and scam accounts from all five social media platforms by individual communication channels.

Channels	Min	Median	Max	Cluster	Accts.	Accts.%
Email	2	3	8	12	44	17.12
Phone	2	3	15	21	88	43.78
URL	2	2	42	41	231	62.60
All (Distinct)	2	3	42	74	355	42.66

The largest cluster contained 42 scam accounts linked through URLs, while the smallest and median cluster sizes across the three communication types were 2 and 3 accounts, respectively.

Key Takeaways. Scammers leverage multiple platforms and interlink accounts to broaden their operations, frequently redirecting users through strategic bio links and aggregating various platforms. Our analysis reveals that platforms such as *YouTube*, *Instagram*, and *Amazon* are often exploited for donation requests. Scammers also use crowdfunding services to solicit both recurring and one-time contributions. Moreover, scammers operate in organized clusters, connecting campaigns through URLs, emails, or phone numbers, showcasing their advanced and coordinated methods for targeting victims.

10 Financial Validation and Tracking Payments

From the profile metadata and post engagements of scammers on five social media platforms, we observed that fraudsters soliciting donations often involve requesting payments via various methods such as *PayPal* and cryptocurrency addresses. To further validate these scams' impact, we partnered with *PayPal*, and *Chainabuse*, sharing 1,898 email addresses with *PayPal*, and 142 cryptocurrency addresses with *Chainabuse*. Below, we present the findings related to these scamming payment profiles based on feedback from our industry partners.

PayPal's Scam Validation. From the 1898 email addresses that were shared, *PayPal* was able to identify and associate 79.71% (1513/1898) of these to *PayPal* accounts on the platform. Among these identified accounts, 26% were restricted at some point during their activity on the platform. Within these 26% restricted accounts, above 50% had more than one restriction placed throughout their time on *PayPal*, and 42% were currently restricted at the time of data sharing. Finally, based on the overall restrictions placed on these accounts, the top reasons were (i) KYC (Know your Customer) & Compliance concerns, and (ii) Risky Operations like Unauthorized Account Access or Creation.

Chainabuse Scam Validation. Out of 142 addresses, 21.83% (31/142) were identified as invalid. We are unclear as to why scammers provide invalid cryptocurrency addresses when soliciting donations. However, we suspect that by using an invalid address, scammers compel victims to contact them for assistance, redirecting the communication in their favor. We provide chain analysis

on the remaining 78.16% (111/142) valid addresses to four popular chains: *Ethereum*, *Binance*, *Polygon*, *Avalance*, and *Bitcoin*; identifying 4 of these as suspicious by these popular chains.

Incoming Volume/Transfer In total, we identified 96 accounts with an average USD value of $2,574,907.09 and a total sum of $247,191,080.45 at the time of writing this paper. Based on the first transfer date of the transaction, we observe that these 75% (72/96) were active first in 2024, and the remaining transactions 25% (24/96) from 2016 to 2013. Scammers using new addresses for transactions are common practices to remain anonymous with the previous transactions history. Among these transactions, we found two long-tail transactions - the first highest recorded account transaction value to $241,251,535 and the second highest was $2,863,122.17. Excluding the first and second highest recorded transactions accounts as long-tail, the remaining 94/96 accounts transactions reflected an average of $32,727.90 and a total sum of $3,076,422.71 value. Among these 96 transactions, we identified 11 transactions valued less than $1, with an average incoming volume of $0.22, and a total sum of $2.41. We suspect these small incoming transactions below $1 are rather an airdropping.

Outgoing Transfers In total, we identified 130 outgoing transfers with an average value of $1,530.04 and a total sum of $198,906.

Disclaimer. Our evaluation is based on the observed transaction histories and reported fraud categories. However, are unable to confirm that all transactions associated with these addresses are connected to scams.

Key Takeaways. Scammers utilize various payment methods, including PayPal and cryptocurrency, to solicit donations while maintaining anonymity. Our collaboration with industry partners reveals that scammer's payment method linked to various fraud topologies including compliance violations and unauthorized activities. We suspect invalid cryptocurrency addresses are used for manipulating victims to pursue direct communication. The cryptocurrency transaction analysis highlights that scammers often use new addresses to obscure histories, while a small number of accounts perform large sums. Although we could not conclude scams involving transactions of $1 or less, we suspect that these may go unnoticed due to small recurring payments or platform monitoring biases. Scammers potentially use small transactions, such as airdrops, which may serve to create plausible activity or evade detection.

11 Dataset Evaluation and Discussion

In this section, we provide details on the evaluation of the dataset through manual inspection. We share observed insights into the limitations and assessed the filtration efficacy of using large language models (GPT-4o) and the reliance on external databases for classifying email addresses, phone numbers, and URLs as malicious along with the studied social media profiles.

Efficacy of LLM-based Filtration. We manually evaluated the effectiveness of using a Large Language Model (LLM) to classify whether a given post is related to a donation context. For this evaluation, we selected 50 posts from each of the social media platforms: *Facebook*, *Instagram*, *Telegram*, *X*, and *YouTube* from both cases, posts that were classified as false and true for donation based context. In total, we manually evaluated 500 posts: 250 from the *True* class and 250 from the *False* class. Our evaluation showed that the LLM achieved 100% efficacy in correctly identifying donation contexts in the *True* class. However, in the *False* class, we observed two main categories where the LLM underperformed: (i) 19/250 posts lacked sufficient donation contextual information, containing only links, emojis, or hashtags with contact details, and (ii) 33/250 posts found in languages other than English, which were classified as *False*. As a result, we suspect that our evaluation might have overestimated false positive cases while maintaining high true positive efficacy. We propose that these limitations can be further addressed by (i) incorporating additional context checks for prevalent hashtags, and inpsecting the landing URL, and (ii) enhancing the LLM's capabilities to better identify donation contexts in languages other than English through multilingual settings.

Reliability of Security Risk Engines. To assess the reliability of the risk engines used to identify malicious URLs, phone numbers, and emails reported under the abuse category, we conducted two distinct evaluations.

The first evaluation involved inspecting potentially malicious URLs from our dataset by manually opening them in a browser. Out of 252 URLs flagged as phishing or malicious, we randomly selected 100 URLs for inspection. Of these, 47 were inactive or taken down. Among the remaining 53 active URLs, 29 were flagged by *Chrome* as potential phishing or malicious sites with a *Deceptive site ahead* warning. Upon visiting these URLs, we found that 14/29 displayed missing content with a default template, while 13 led to fake donation pages for various causes, such as child support, healthcare, and relief, and 2 were redirected to sign-up pages without further information. For the other 24 active URLs, although they were marked as malicious by the *VirusTotal API*, no deceptive banner was shown upon visiting. However, upon further inspection, each of these 13 URLs was missing content or had been removed, and 11 consisted of solicitations for donations through sign-up or payment information submission pages. For each of these 13 URLs, we found that 1/68 vendors on *VirusTotal* flagged them as suspicious or malicious, while the responses from 68 other vendors were marked as clean. Since phishing sites are often ephemeral and missing content makes classification challenging, not all vendors may have processed these URLs promptly before the content change. We suggest that such cases could be improved through regular monitoring and by consolidating responses from multiple vendors to enhance URL flagging accuracy.

In the second evaluation of phone numbers and email addresses, we conducted additional analyses using two datasets: (i) 50 known malicious entries (25 phone numbers and 25 email addresses) from publicly reported corpus [84], and (ii) a benign dataset of 50 entries from the authors' friends and family (25 phone numbers and 25 email addresses). We queried these 100 entries against the risk engine and found that 19/25 phone numbers and 23/25 email addresses were flagged with risk levels above 85%. However, 6 phone numbers and 2 email addresses showed risk percentages between

5% and 65%, making them unreliable for classification as malicious. In contrast, all 50 entries from the benign dataset were marked with 0% risk. Although the risk engine performed inconsistently for email and phone number assessments with lower risk percentages for known corpus, we argue that integrating multiple providers and combining scores could potentially enhance results which would require additional resources.

Social Media Profiles and Scam Prevalence. We randomly selected 100 social media accounts from the dataset and manually inspected them using a browser. Our findings revealed that 9/100 accounts had been deactivated by the social media platforms for violating terms and conditions, and 17/100 were either deactivated or deleted by the users. For the remaining 74/100 active accounts, we manually reviewed their public profiles and engagement. Of these, 14 accounts displayed default profile pictures and had limited public interaction, while 18 accounts were used solely for retweets and shares, with no original posts. We suspect that these accounts are used to harvest followers or create the appearance of an organically aged social media profile. The remaining 42/74 accounts were found to engage in some form of donation solicitation, targeting causes such as ongoing war and human welfare programs (18 accounts), education and local training programs (11 accounts), local wildlife foundations seeking donations for preservation efforts (6 accounts), single mom and women support (3 accounts), and other miscellaneous disadvantaged groups (4 accounts).

12 Recommendations

Based on our observations and findings, we propose recommendations to combat donation-based abuses. These recommendations are intended for adoption by social media platforms, financial services, crowdfunding platforms, and platform users. We provide further details below.

Recommendations to Social Media Platforms. We suggest that social media platforms adopt a detection measurement setup similar to the one proposed in our research. For proactive prevention, social media platforms can utilize a fraud score to assess whether the email address or phone number used during sign-up poses a fraud risk. Similarly, for reactive measures against existing profiles, we recommend monitoring the use of external media associated with profile bio-data or shared posts. Our network analysis of donation abuse revealed that scammers often operate across multiple social media platforms as part of their modus operandi. We encourage social media platforms to share information with other platforms about detected suspicious behaviors to prevent such fraudulent activities. Implementing a warning message for regular users when a social media post contains donation requests from flagged cryptocurrency addresses or payment links could help users avoid potential interactions with scammers.

Recommendations to Financial In-Take Services. We recommend that financial intake services, specifically crowdfunding platforms and payment profiles, monitor the URLs shared across their platforms. For instance, crowdfunding services like *GoFundMe*, *Fundly*, *PayPal*, and others often include links that scammers use to request payouts. These financial intake services can effectively implement referral header monitoring techniques based on the source of visits. Referral headers contain links and source information indicating where a user is directed from. By monitoring referral headers and assessing whether a social media profile is linked to fraudulent activity, crowdfunding platforms, and payment services can reduce the risk of funding abuse by scammers.

Recommendations to Social Media Users. We recommend social media users conduct thorough fact-checking before supporting any donation-related efforts. This includes verifying bio data, and affiliations, understanding the purpose and planned use of funds, and reviewing feedback from other donors. For instance, databases tracking charity affiliations are valuable resources for authenticating charitable organizations. When donating to individuals or private causes, we recommend users support only when there is a known connection and look out for any account duplications or impersonations.

Key Takeaways. We provide recommendations to combat donation-based abuses on social media platforms, financial services, crowdfunding platforms, and among users. Social media platforms are encouraged to adopt fraud scores for proactive detection and monitor external media for suspicious activity. Financial services are suggested to monitor URLs for scams and use referral header monitoring to reduce fraud risks. Users are urged to conduct thorough checks before donating, verify affiliations, and exercise caution, particularly when supporting unfamiliar causes or individuals.

13 Conclusion

In this research, we presented the first large-scale study of donation-based abuses across five social media platforms: *X*, *Instagram*, *Facebook*, *Telegram*, and *YouTube*. By analyzing data from over 150K social media users and 3 million posts, we identified over 832 scammers soliciting fraudulent donations on these platforms. Our analysis of profile creation and user engagement revealed scammers' techniques for luring victims and requesting payments through payment profiles such as *PayPal*, cryptocurrency addresses, crowdfunding services, and survey forms. Our measurement approach identified scam accounts operating on 11 platforms (9 social media, and 2 e-commerce) beyond their origins. Through collaboration with industry partners *PayPal* and the cryptocurrency abuse database *Chainabuse*, we validated the scams and assessed the financial impact of these fraudulent accounts. Furthermore, we provided detailed disclosures to affected entities and proposed recommendations to protect against future abuses.

Acknowledgment

We sincerely thank Ian Schade from Chainabuse for sharing valuable insights regarding cryptocurrency accounts. Our appreciation also goes to Qutub Khan Asghar Vajihi from PayPal for providing insights related to PayPal accounts. Additionally, we are grateful to Muhammad Saad from X (formerly Twitter) for his initial discussions on donation-based scams prevalent in the contexts of the X platform. This work was funded by the German Federal Ministry of Education and Research (BMBF grant 16KIS1900 "UbiTrans"); and by the EU—NGEU National Sustainable Mobility Center (CN00000023), Italian Ministry of University and Research Decree

n. 1033—17/06/2022 (Spoke 10). Lastly, this work was carried out while Dario Lazzaro was enrolled in the Italian National Doctorate on Artificial Intelligence run by the Sapienza University of Rome in collaboration with the University of Genoa.

References

[1] FTC, "Charity fraud." https://consumer.ftc.gov/features/pass-it-on/charity-fraud.

[2] FTC, "Scam 'charities' will take your money and run." https://www.fcc.gov/scam-charities-will-take-your-money-and-run.

[3] FBI, "Charity and disaster fraud." https://www.fbi.gov/how-we-can-help-you/scams-and-safety/common-scams-and-crimes/charity-and-disaster-fraud.

[4] F. C. M. East, "Fake donation emails and websites rise amid the israel-hamas war." https://fastcompanyme.com/news/fake-donation-emails-and-websites-rise-amid-the-israel-hamas-war/, 2023.

[5] M. L. Gutzwiller, "Spotting charity scams: How to give safely." https://www.cshco.com/articles/spotting-charity-scams/, 2024.

[6] D. Amato, "Text and email scams to watch for in 2024." https://www.rbcroyalbank.com/en-ca/my-money-matters/money-academy/cyber-security/understanding-cyber-security/text-and-email-scams-to-watch-for-in-2024/, 2024.

[7] N. C. C. Council, "Charity donation fraud." https://newcastle.gov.uk/services/business-and-commerce/business-commerce/trading-standards/campaigns/charity-donation-fraud.

[8] C. Boyd, "Beware of fake twitter philanthropists offering to put $750 into your cash app account." https://www.malwarebytes.com/blog/news/2022/04/beware-of-fake-twitter-philanthropists-offering-750-for-your-cash-app-account.

[9] C. POPOV, "Charity scams: How to spot and avoid fake charities." https://www.bitdefender.com/blog/hotforsecurity/protect-your-donations-spot-and-avoid-fake-charities/.

[10] M. Keane, "Charity fraud." https://www.britannica.com/money/charity-fraud.

[11] CAF, "Why charities should be cyber-aware." https://www.cafonline.org/about-us/security-centre/be-aware---current-threats/scam-emails.

[12] I. Support, "Scam alert: "donation to charity or prize winning"." https://www.uregina.ca/is/security/advisories/security-advisory56.html, 2023.

[13] B. Dean, "Social network usage & growth statistics: How many people use social media in 2024?." https://backlinko.com/social-media-users, 2024.

[14] B. Matthews, "Social media stats for charities and nonprofits." https://empower.agency/social-media-stats-charities-nonprofits/.

[15] J. Tabas, "How nonprofit organizations can boost donations via social media." https://www.forbes.com/sites/allbusiness/2024/05/10/how-nonprofit-organizations-can-boost-donations-via-social-media/, 2024.

[16] C. Water, "Protect yourself from charitable-giving scams." https://clearwatercreditunion.org/protect-yourself-from-charitable-giving-scams-2023-12-04/, 2023.

[17] AT&T, "Social media charity scam." https://about.att.com/pages/cyberaware/ar/social-media-charity-scam.

[18] E. Fletcher, "Social media: a golden goose for scammers." https://www.ftc.gov/news-events/data-visualizations/data-spotlight/2023/10/social-media-golden-goose-scammers, 2023.

[19] IRS, "Dirty dozen: Irs warns of scammers using fake charities to exploit taxpayers." https://www.irs.gov/newsroom/dirty-dozen-irs-warns-of-scammers-using-fake-charities-to-exploit-taxpayers, 2023.

[20] CNBC, "Fake charities can be almost impossible to spot. here's how to make sure your donations get to the right place." https://www.cnbc.com/2022/07/07/how-to-avoid-charity-impersonation-scams-in-times-of-crisis.html, 2022.

[21] G. Torre, "Warning issued over fake social media accounts running flood donation scams." https://nit.com.au/17-01-2023/4743/warning-issued-over-fake-social-media-accounts-running-flood-donation-scams, 2023.

[22] S. Watch, "Scam statistics." https://www.scamwatch.gov.au/research-and-resources/scam-statistics?scamid=14&date=2024, 2024.

[23] E. News, "Charity warns 2023 was the worst year for child sexual abuse content." https://www.euronews.com/next/2024/04/23/german-internet-domain-used-by-criminal-groups-in-worst-year-for-online-child-sexual-abuse, 2024.

[24] Z. Ali, "The rise of ai is creating a rise in scams on social media." https://www.howtogeek.com/the-rise-of-ai-is-creating-a-rise-in-scams-on-social-media/, 2024.

[25] T. Riley, "Cybercriminals are posing as ukraine fundraisers to steal cryptocurrency." https://cyberscoop.com/cybercriminals-are-posing-as-ukraine-fundraisers-to-steal-cryptocurrency/, 2022.

[26] M. Britton, "Attackers exploit middle east crisis to solicit fraudulent cryptocurrency donations for children." https://abnormalsecurity.com/blog/attackers-exploit-middle-east-crisis-solicit-cryptocurrency-donations, 2023.

[27] NZU, "Cybercriminals abuse advertisement on x to promote crypto scam." https://news.zke.com/cybercriminals-abuse-advertisement-on-x-to-promote-crypto-scam/, 2024.

[28] NZU, "Cybercriminals abuse advertisement on x to promote crypto scam." https://news.zke.com/cybercriminals-abuse-advertisement-on-x-to-promote-crypto-scam/, 2024.

[29] G. Chow, "Trumped up crypto scams – criminals deploy trump donation scams." https://www.netcraft.com/blog/trumped-up-crypto-donation-scams/, 2024.

[30] S. D. Github, "Code/data share." https://github.com/CISPA-SysSec/scam_donation, 2024.

[31] A. Oest, Y. Safaei, A. Doupé, G.-J. Ahn, B. Wardman, and K. Tyers, "Phishfarm: A scalable framework for measuring the effectiveness of evasion techniques against browser phishing blacklists," in *2019 IEEE Symposium on Security and Privacy (SP)*, 2019.

[32] A. Oest, P. Zhang, B. Wardman, E. Nunes, J. Burgis, A. Zand, K. Thomas, A. Doupé, and G.-J. Ahn, "Sunrise to sunset: Analyzing the end-to-end life cycle and effectiveness of phishing attacks at scale," in *USENIX Security*, 2020.

[33] B. Acharya and P. Vadrevu, "{PhishPrint}: Evading phishing detection crawlers by prior profiling," in *USENIX Security*, 2021.

[34] P. Peng, L. Yang, L. Song, and G. Wang, "Opening the blackbox of virustotal: Analyzing online phishing scan engines," in *ACM Internet Measurement Conference (IMC)*, 2019.

[35] P. Zhang, A. Oest, H. Cho, Z. Sun, R. Johnson, B. Wardman, S. Sarker, A. Kapravelos, T. Bao, R. Wang, Y. Shoshitaishvili, A. Doupé, and G.-J. Ahn, "Crawlphish: Large-scale analysis of client-side cloaking techniques in phishing," in *IEEE Security and Privacy (IEEE S&P)*, 2021.

[36] K. Subramani, W. Melicher, O. Starov, P. Vadrevu, and R. Perdisci, "Phishinpatterns: measuring elicited user interactions at scale on phishing websites," in *ACM Internet Measurement Conference (IMC)*, 2022.

[37] J. Liu, P. Pun, P. Vadrevu, and R. Perdisci, "Understanding, measuring, and detecting modern technical support scams," in *IEEE European Symposium on Security and Privacy (Euro S&P)*, 2023.

[38] N. Miramirkhani, O. Starov, and N. Nikiforakis, "Dial one for scam: A large-scale analysis of technical support scams," in *Network and Distributed System Security Symposium (NDSS)*, 2017.

[39] T. Liu, Y. Zhang, J. Shi, Y. Jing, Q. Li, and L. Guo, "Towards quantifying visual similarity of domain names for combating typosquatting abuse," in *IEEE Military Communications*, 2016.

[40] N. Nikiforakis, M. Balduzzi, L. Desmet, F. Piessens, and W. Joosen, "Soundsquatting: Uncovering the use of homophones in domain squatting," in *Information Security International Conference (ISC)*, 2014.

[41] N. Nikiforakis, S. Van Acker, W. Meert, L. Desmet, F. Piessens, and W. Joosen, "Bitsquatting: Exploiting bit-flips for fun, or profit?," in *World Wide Web (WWW)*, 2013.

[42] P. Agten, W. Joosen, F. Piessens, and N. Nikiforakis, "Seven months' worth of mistakes: A longitudinal study of typosquatting abuse," in *Symposium on Network and Distributed System Security (NDSS)*, 2015.

[43] Y.-M. Wang, D. Beck, J. Wang, C. Verbowski, and B. Daniels, "Strider typo-patrol: Discovery and analysis of systematic typo-squatting.," in *USENIX Security*, 2006.

[44] J. Szurdi, B. Kocso, G. Cseh, J. Spring, M. Felegyhazi, and C. Kanich, "The long taile of typosquatting domain names," in *USENIX Security*, 2014.

[45] P. Vadrevu and R. Perdisci, "What you see is not what you get: Discovering and tracking social engineering attack campaigns," in *ACM Internet Measurement Conference(IMC)*, 2019.

[46] A. Zarras, A. Kapravelos, G. Stringhini, T. Holz, C. Kruegel, and G. Vigna, "The dark alleys of madison avenue: Understanding malicious advertisements," in *ACM Internet Measurement Conference (IMC)*, 2014.

[47] B. Srinivasan, A. Kountouras, N. Miramirkhani, M. Alam, N. Nikiforakis, M. Antonakakis, and M. Ahamad, "Exposing search and advertisement abuse tactics and infrastructure of technical support scammers," in *World Wide Web (WWW)*, 2018.

[48] X. Li, A. Yepuri, and N. Nikiforakis, "Double and nothing: Understanding and detecting cryptocurrency giveaway scams," in *Network and Distributed Systems Security (NDSS)*, 2023.

[49] P. Xia, H. Wang, X. Luo, L. Wu, Y. Zhou, G. Bai, G. Xu, G. Huang, and X. Liu, "Don't fish in troubled waters! characterizing coronavirus-themed cryptocurrency scams," in *APWG Symposium on Electronic Crime Research (eCrime)*, 2020.

[50] R. Phillips and H. Wilder, "Tracing cryptocurrency scams: Clustering replicated advance-fee and phishing websites," in *IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, 2020.

[51] G. Hong, Z. Yang, S. Yang, X. Liao, X. Du, M. Yang, and H. Duan, "Analyzing ground-truth data of mobile gambling scams," in *IEEE Symposium on Security and Privacy (IEEE S&P)*, 2021.

[52] B. Acharya, M. Saad, A. E. Cinà, L. Schönherr, H. D. Nguyen, A. Oest, P. Vadrevu, and T. Holz, "Conning the crypto conman: End-to-end analysis of cryptocurrency-based technical support scams," *IEEE Security and Privacy (IEEE S&P)*, 2024.

[53] X. Li, A. Rahmati, and N. Nikiforakis, "Like, Comment, Get Scammed: Characterizing Comment Scams on Media Platforms," in *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2024.

[54] M. Mirtaheri, S. Abu-El-Haija, F. Morstatter, G. Ver Steeg, and A. Galstyan, "Identifying and analyzing cryptocurrency manipulations in social media," in *IEEE Transactions on Computational Social Systems (ITCSS)*, 2021.

[55] S. Khaled, N. El-Tazi, and H. M. Mokhtar, "Detecting fake accounts on social media," in *IEEE International Conference on Big Data (ICBG)*, 2018.
[56] J. Mink, L. Luo, N. M. Barbosa, O. Figueira, Y. Wang, and G. Wang, "{DeepPhish}: Understanding user trust towards artificially generated profiles in online social networks," in *USENIX Security*, 2022.
[57] B. Acharya, D. Lazzaro, E. López-Morales, A. Oest, M. Saad, A. Emanuele Cinà, L. Schönherr, and T. Holz, "The imitation game: Exploring brand impersonation attacks on social media platforms," in *USENIX Security*, 2024.
[58] X. Li, A. Rahmati, and N. Nikiforakis, "Like, comment, get scammed: Characterizing comment scams on media platforms," *Network and Distributed System Security Symposium (NDSS)*, 2024.
[59] M. T. Whitty, "Is there a scam for everyone? psychologically profiling cyberscam victims," *European Journal on Criminal Policy and Research*, 2020.
[60] J. S. Albanese, "Fraud: The characteristic crime of the twenty-first century," *Trends in Organized Crime*, 2005.
[61] A. A. Gillespie and S. Magor, "Tackling online fraud," in *ERA Forum*, Springer, 2020.
[62] S. Wood, D. Hengerer, and Y. Hanoch, "Scams in the time of covid-19:: Pandemic trends in scams and fraud," in *A Fresh Look at Fraud*, pp. 42–57, Routledge, 2022.
[63] Twitter, "User detail twitter api." https://developer.twitter.com/en/docs/twitter-api/v1/accounts-and-users/follow-search-get-users/api-reference/get-users-lookup, 2024.
[64] Twitter, "User timelines twitter api." https://developer.twitter.com/en/docs/twitter-api/tweets/timelines/introduction, 2024.
[65] Apify, "Apify instagram scraper api." https://apify.com/apify/instagram-scraper, 2024.
[66] D. Milevski, "Apify telegram scraper api." https://apify.com/danielmilevski9/telegram-channel-scraper, 2024.
[67] D. Milevski, "Telemetrio telegram scraper api." https://telemetr.io/, 2024.
[68] Apify, "Youtube scraper." https://apify.com/streamers/youtube-scraper, 2024.
[69] Apify, "Facebook scraper." https://apify.com/apify/facebook-posts-scraper, 2024.
[70] O. Platform, "Models - openai api (gpt-4o)." https://platform.openai.com/docs/models/gpt-4o.
[71] VirusTotal, "Virustotal api v3 overview." https://docs.virustotal.com/reference/overview.
[72] I. E. V. API, "Email validation api documentation." https://www.ipqualityscore.com/documentation/email-validation-api/overview.
[73] I. P. V. API, "Phone validation api documentation." https://www.ipqualityscore.com/documentation/phone-number-validation-api/overview.
[74] G. Bowyer, "CLD2-CFFI – Python (CFFI) Bindings for Compact Language Detector 2," 2016. https://github.com/GregBowyer/cld2-cffi.
[75] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," in *Empirical Methods in Natural Language Processing (EMNLP)*, 2019.
[76] M. Grootendorst, "Bertopic: Neural topic modeling with a class-based tf-idf procedure," *arXiv preprint arXiv:2203.05794*, 2022.
[77] L. McInnes, J. Healy, N. Saul, and L. Großberger, "Umap: Uniform manifold approximation and projection," *Journal of Open Source Software*, 2018.
[78] L. McInnes, J. Healy, and S. Astels, "hdbscan: Hierarchical density based clustering," *Journal of Open Source Softw.*, 2017.
[79] M. Grootendorst, "Keybert: Minimal keyword extraction with bert.." https://doi.org/10.5281/zenodo.4461265, 2020.
[80] K. R. Shahapure and C. K. Nicholas, "Cluster quality analysis using silhouette score," *Data Science and Advanced Analytics (DSAA)*, 2020.
[81] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," in *International Conference on Machine Learning (ICML)*, 2021.
[82] T. Hastie, J. H. Friedman, and R. Tibshirani, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, 2001.
[83] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," in *IEEE Transactions on Pattern Analysis and Machine Intelligence (ITPAMI)*, 2002.
[84] ScamSearch, "Global scam database." https://scamsearch.io/.

A Prompt Engineering on Donation Context

In this section, we provide details on creating prompt injection in identifying the posts that are related to the donation context. We chose LLMs specifically for their effectiveness and adaptability in handling diverse natural language processing tasks, making them ideal for accurately classifying fraudulent donation solicitations.

To determine if a post is related to donation solicitations, we designed a prompt that evaluates whether the input post includes donation requests, outputting the result as a boolean (true or false). Using the OpenAI API [70], we queried posts from the five social media platforms to obtain their respective outputs. Below, we provide examples of prompt instruction along with input samples for responses received in both cases (false and true).

Prompt Instruction.

> You are given a text and must identify whether it is requesting money, donations, or charity support. The output should be a boolean value compatible with a Python boolean value. Do not include any explanation.

Input Sample Post - API Response True Case.

> WE JUST HIT OUR GOAL OF $500 of donations to Extra Life. We would like to thank everyone who donated to this great cause!

Output of ChatGPT - API Response True Case.

> True

Input Sample Post - API Response False Case.

> RT @bbby**luve: Oi meus amores! We are only 15 days away from Brazil fanmeeting? Are you ready for that amazing night??

Output of ChatGPT - API Response False Case.

> False

Figure 6: Visualization of 50 random samples from *Association-Logos* (left), *Games-Cartoon* (right) clusters of scammers.

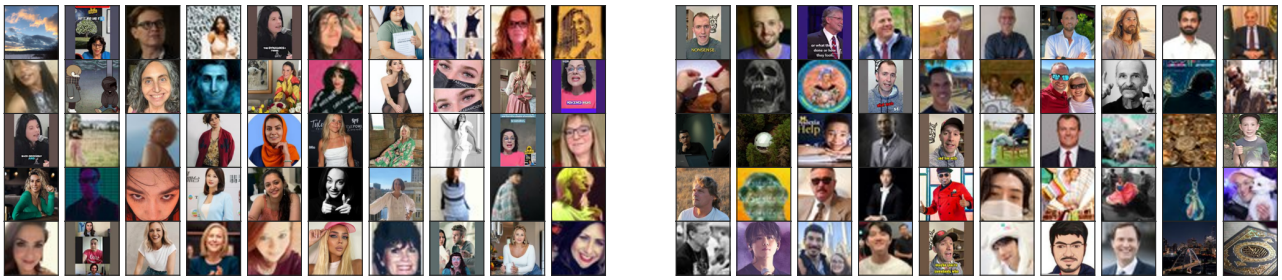


Figure 7: Visualization of 50 random samples from *Female* (left) *Male* (right) clusters of scammers.

Figure 8: Visualization of 50 random samples from *Politics-War* (Left), and *Pets* (right) clusters of scammers.

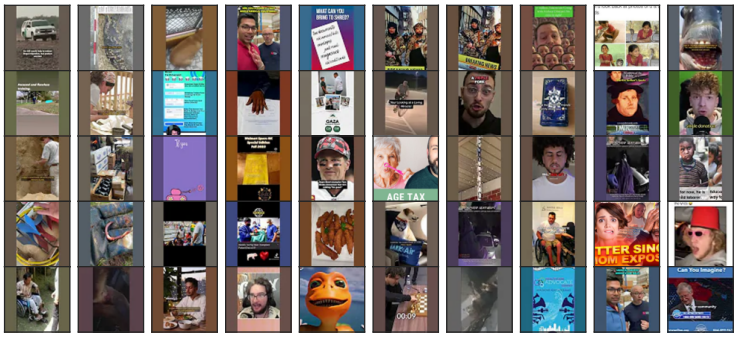


Figure 9: Visualization of 50 random samples from Miscellaneous clusters of scammers.