“You Have Been Selected as the Winner”: Characterizing User-Reported Scams on TikTok

Smirity Kaushik [*]
The George Washington University

Kyle Beadle
University College London

Gauri Nayak
Cornell University

Madelyn Rose Sanfilippo
University of Illinois Urbana-Champaign

Mainack Mondal
IIT Kharagpur

Yang Wang
University of Illinois Urbana-Champaign

Sai Teja Peddinti
Google

Jingjie Li
University of Edinburgh

Yixin Zou
Max Planck Institute for Security and Privacy

Abstract

Short-form video platforms (SVPs) such as TikTok have grown rapidly in popularity. While online scams have been extensively studied, the extent to which they take new forms on SVPs and the discourses around them remain understudied. Using TikTok as a case study, we analyzed 150 videos in which content creators reported scam experiences and offered anti-scam advice. We focus on how TikTok users (creators, followers, and commenters) discuss scams, rather than analyzing scams. Our analysis surfaces six types of scams, including creator impersonation and account badge verification scams that target TikTok's influencer-follower ecosystem. Scammers also exploit platform-specific features (e.g., direct messaging or the “For You Page”) to lure victims. In response, TikTok users share strategies to identify scammer profiles and communication cues, building community support through anti-scam advice. Based on our findings, we offer recommendations for systemizing platform support to combat scams and leveraging the influencer ecosystem to raise awareness.

1 Introduction

Fraudulent content on social media is on a rapid rise. According to the US Federal Trade Commission (FTC), consumers lost up to $3 billion in online scams in 2024, 70% of which started through social media [39]. Short-form video platforms (SVPs), such as TikTok, Instagram, and YouTube Shorts, are gaining dominance in the social media market, yet they are not immune to the risks of fraudulent content. Online scams have been studied extensively [36] as researchers examine pig-butchering scams [4, 71], romance scams [114], and crypto scams [117] by analyzing victim testimonies [70, 71, 73] or measuring actual scam operations [5, 89]. However, SVPs are relatively new and offer distinct affordances, notably their bite-sized video formats and influencer-centric culture [6, 13, 67], that may reshape traditional scams and amplify their risks. Current SVP literature has largely focused on documenting scam operations, such as malicious livestream selling strategies [116] or the illicit trade of social media accounts [10]. While emerging research has begun to explore the human element of SVPs, e.g., anti-security advice [111] and social surveillance [49], a critical gap remains: it is still unclear how users encounter, report, and react to scams on SVPs.

In this paper, we study user-reported scams on SVPs using TikTok as a case study. We use the term *TikTok users* to refer to content creators, their followers, and viewers/commenters on their videos. We focus on TikTok for three reasons: (i) *Scale*: With more than 1.5 billion users worldwide—over half of whom are under 35 years old [27]—TikTok is among the most popular SVPs. (ii) *Culture*: TikTok's short-form video format, combined with algorithm-driven recommendation, has fostered an influencer-centric and commerce-driven culture [6, 13, 67]. TikTok has emerged as a platform for e-commerce and online advertising, benefiting 7 million businesses and over 100 million consumers in the U.S. [95], while also creating incentives for scams. Simultaneously, it is also a rich site for examining user experiences, beliefs, and advice across diverse topics including security advice [111] and interpersonal surveillance [49]. (iii) *API access*: TikTok allows researchers to access and study public TikTok content and account data through its *Research Tools* program [97]. In contrast, YouTube's *Data API* does not isolate short-form content, and Instagram's *Meta Content Library and API* does not include machine-generated video transcripts, limiting utility.

We qualitatively analyze 150 English-language TikTok videos from the US, where TikTok users report scams experi-

Copyright is held by the author/owner. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee.
USENIX Symposium on Usable Privacy and Security (SOUPS) 2026.
August 23–26, 2026, Hannover, Germany.

[*]This work was conducted while the author was a visiting PhD Scholar at the Max Planck Institute for Security and Privacy.

enced on the platform or elsewhere, and/or share anti-scam advice. We identified relevant videos through hashtag-based keyword search, applied filtering criteria, and conducted thematic analysis to distill key themes from the videos. We also conducted sentiment analysis of 1,163 associated comments and analyzed video/account metadata to validate the credibility of videos and their creators. Our findings address the following research questions:

RQ1 *What types of scams do TikTok users report?*

We synthesized user-reported scams into six categories. While these categories are broadly aligned with prior literature [9, 36], some of them evolved in distinct ways on TikTok. Additionally, we identified two new types of scams. Overall, we found that platform-specific affordances and influencer-driven culture [57] shape how scams are experienced and the harms they caused. E.g., impersonation and account badge verification scams specifically target influencers and their followers, exploiting trust built within communities.

RQ2 *How do user-reported scams manifest on TikTok?*

TikTok users shared that scammers approached them through TikTok-specific features, such as direct messaging and the 'For You Page' (FYP) [103]. The luring techniques used are grounded in established psychological principles [90], but are adapted to increase the effectiveness on the TikTok platform. The scam entry points and escalation patterns also vary significantly by scam type.

RQ3 *How do TikTok users identify and redress these scams?*

TikTok users reported various signals they leveraged to identify scams, and some of these are very specific to TikTok's culture and communication norms. These TikTok-specific signals focus on scammer communication patterns, e.g., use of templated messages or unconventional language, or analysis of inconsistencies between authentic creator profiles and impersonated accounts. Users also shared anti-scam advice and redressal mechanisms, often grounded in personal experience, that we categorize and map to the scam types.

By analyzing user-reported scams on TikTok, we make four contributions. First, our work extends existing scam taxonomies to the SVP context by highlighting how platform affordances and the influencer-centric culture reshape scam experiences. Second, we report how scammers leverage TikTok-specific features and adapt luring techniques to increase the effectiveness of scams on the platform. Third, our work shows that while perpetuating online scams, TikTok also serves as a peer-support space where users actively share pro-security advice and community-driven strategies to prevent and mitigate scams, in contrast to prior SOUPS literature that characterizes TikTok as the breeding ground for anti-security advice and interpersonal surveillance [49, 111]. Lastly, we identify implications for SVPs, including privacy-preserving on-device scam detection with real-time warnings, and leveraging influencer and multi-channel network brand manager ecosystems to deliver large-scale anti-scam education.

2 Background and Related Work

2.1 Online Scams

Types of scams While the definitions of scams[1] vary in academic literature, we take the following working definition from Beals et al. [9]—the scam or fraud must "intentionally and knowingly deceive the victim by misrepresenting, concealing, or omitting facts about promised goods, services, or other benefits . . . for the purpose of monetary gain." There are several existing taxonomies of scams. Beals et al.'s taxonomy focuses on financial fraud, including nine different fraud categories (e.g., consumer investment fraud, employment fraud, and charity fraud) across both digital and in-person channels [9]. Federal Trade Commission (FTC) also categorizes fraud based on consumer reports [36]. Other scholars examined specific scam types such as online shopping [54, 66], pig-butchering [71], romance [114], deepfake-enabled fraud [112, 118], and cryptocurrency scams [117], including analyses of victim testimonies [70, 73] and large-scale investigations of scam operations [5, 89]. We build on these prior works to explore the different types of SVP scams.

Scammer tactics Stajano and Wilson summarized seven principles scammers typically leverage to lure victims: distraction, social compliance, herd, dishonesty, kindness, need and greed, and time [90]. Prior research in phishing and social engineering has extensively validated these principles across various digital domains. For instance, Ferreira et al. examined how principles of persuasion are operationalized in phishing attacks [40]. Furthermore, Salahdine et al. examined how social engineering techniques are used to manipulate individuals into performing actions or divulging confidential information, and how modern deceptive practices frequently use hybrid methods that exploit both human psychology and platform-specific vulnerabilities [83]. Other studies have characterized scammers' tactics and operations with regard to particular types of scams [4, 61, 71, 87, 106].

Button et al. provided insights into various reasons why victims fall for scams [19]. Several studies have highlighted the demographics and psychological characteristics of scam victims, such as how romance scam victims tend to be middle-aged, well-educated women who are more impulsive and trustworthy [113], and scam victims tend to have less financial knowledge [56] and lower cognitive reflectiveness [85].

In this work, we adopt the lure principles from Stajano et al. [90] and explore tactics used by scammers on TikTok,

[1] Prior literature has often used the terms "scam" and "fraud" interchangeably [9, 16] and we do this in our paper too.

including how they leverage TikTok's influencer culture and product features (direct messages and 'For You Page').

2.2 Security Folklore and Advice

Folklore encompasses people's attitudes, feelings, and beliefs communicated through word of mouth or examples [29]. Security folklore, specifically, is "a collection of beliefs about security practices that have become part of the culture and are transmitted informally through, for example, verbal tales, written social media posts, or demonstrated behavior" [34]. Prior work has studied security folklore by analyzing stories and lessons people share about security [75, 78] and by exploring users' mental models of security technology [77,107,115]. Security folklore focuses on beliefs despite the lack of empirical evidence.

On top of a large body of related work on security advice [33,81], prior work has specifically highlighted the role of peer influence in security behavior adoption [24, 108]. Other work has shown that users go to social media to express critique and frustration about existing password practices [30]—even create jokes about particular advice [11]—or spread anti-security and anti-privacy advice [111]. Influencers, or social media creators with a large audience, could be a powerful source to spread security and privacy advice [93], similar to how they function in brand marketing [51].

In our study, we analyze user-reported scams as a type of security folklore and investigate the role of TikTok creators in the sharing of security and privacy advice.

2.3 Background and Related Work on TikTok

TikTok's rapid rise in popularity has prompted the creation of other short-form video platforms such as Instagram Reels, YouTube Shorts, and Douyin (the Chinese equivalent to TikTok). Content creators frequently post the same videos across many different SVPs [86]. Users' sensemaking of TikTok's highly personalized, black-box algorithm has been the focus of prior research [12,25,26,58,91]. Other work explored how users interact with TikTok to enable activism [21, 53, 60] and to find emotional support [7, 65].

Security and privacy on TikTok Prior work has explored how online harms transpire through TikTok, including spreading disinformation or misinformation [88], or spreading memes as propaganda during Russia's invasion of Ukraine [14]. Other work investigates the harms experienced by marginalized and vulnerable populations, such as Black content creators [52], BlindTokers [63], and transfeminine creators [26]. Focusing on the advice available to users on SVPs, Elgedawy et al. examined the differences in security advice served to children and adults [31], Wei et al. explored the spread of anti-privacy and anti-security advice on TikTok [111], and Gilbert et al. [49] analyzed social surveillance on TikTok.

Online social media accounts, including TikTok, are being abused to conduct fraud and scams [10, 116]. This exploitation extends to the influencer-centric culture of these platforms, where creators frequently engage in brand collaborations and user-generated content (UGC) contracts. In such arrangements, vendors compensate creators for producing sponsored promotional content, often in exchange for product endorsements or demonstrations [3, 62]. While these collaborations are a central part of influencer-driven marketing, scholars have noted that influencer labor is often characterized by informality and power asymmetries between creators and vendors [28, 76]. These collaborations can create opportunities for scammers to exploit creators under the guise of legitimate brand partnerships [48]. With the growing concerns about fraudulent activity on the platform, in 2023 the U.S. FTC ordered TikTok to disclose how it manages fraudulent advertising and scams [47]. Our work complements and extends this prior research by exploring different types of scams reported or discussed on the TikTok platform, the techniques scammers use to defraud users, and how users identify and mitigate these scams.

3 Methodology

We collected and analyzed 150 videos in English posted by TikTok users in the U.S., including three types: reporting a scam (108/72%), providing scam-related advice (15/10%), and doing both (27/18%). We further discuss the ethics and limitations of our approach.

3.1 Data Collection

Identify relevant videos We systematically curated and refined hashtags to identify videos relevant to our research questions, similar to prior work on TikTok [49,111]. We began by exploring videos on TikTok using the hashtag `#scams` to find videos where creators reported experiences with scams. Our initial search yielded 22 relevant videos and led us to find other possible hashtags, such as `#onlineshoppingscam`, `#romancescam`, `#tiktokshopscam`, and `#dminstagram`. We then explored additional videos associated with these hashtags and identified three hashtags that consistently appeared in videos tagged by creators who were reporting personal scam experiences or shared scam-related advice: `#socialmediascam`, `#tiktokscam`, and `#instagramscam`. Using these three hashtags, we queried the TikTok US Researcher API[2] with a daily limit of 100 records, between January and May 2024, yielding 288 videos. No additional filters

[2] We used the official TikTok API in line with the platform's policy, as TikTok restricts using external web crawlers to scrape videos. Since the API is tied to the lead author's primary institution, it allows the collection of videos only in the US region.

were applied. We selected the random selection parameter, meaning the process is not subject to trending or personalized content via TikTok's recommender system. The API searches TikTok's internal index of public videos and returns matching videos with metadata related to the video's engagement and exposure [3]. Because of the 100-per-request cap, we don't know what fraction of videos are returned and cannot claim representativeness.

To ensure the comprehensiveness of our video search, we further experimented with additional hashtags. From the 288 videos, we extracted over 300 additional hashtags such as `#bewareofscammers`, `#scammerredemption`, and `#scamalert`.[4] Using the expanded set of hashtags, we collected 1,272 videos from July to December 2023. Applying the inclusion and exclusion criteria (§3.1), we manually reviewed a random 10% of the sample and found 92% were off-topic. The additional hashtags did not help surface relevant videos but rather introduced substantial noise by pulling in off-topic videos. E.g., a creator used `#scamalert` when comparing the real estate market before the 2008 financial crisis to the present-day market. So we leveraged 288 videos identified by three hashtags: `#socialmediascam`, `#tiktokscam`, and `#instagramscam` for further filtering.

Filter videos We decided on the filtering criteria through iterative discussions among the research team. First, the video must be relevant to scams or fraud, as defined in prior work [9, 22], i.e., it must involve an *intentional* deception by the scammer to distinguish it from other unfair business practices. We also viewed impersonation within scope if it is done for monetary gain, since it constitutes a form of deception. Second, for videos that report scams, the creator must experience a scam from another person (such as a vendor or another creator), and the creator should present evidence that the scam has occurred and any potential consequences (e.g., in terms of monetary harm and emotional distress). Third, for videos that share anti-scam advice, the creator should draw on a specific incident they experienced personally or on someone else's, or offer generic advice on staying safe from scams.

Based on the criteria above, two researchers went through the 288 videos together and conducted the filtering. We resolved any disagreements until consensus was reached [20], leading to 150 videos in our final dataset. Our analysis yielded a good inter-coder reliability (Cohen's Kappa = 0.85) [43]. Examples of excluded videos include: (1) content irrelevant to scams based on prior work's conceptual definitions; (2) videos that mention scams but do not provide sufficient details to indicate a scam has occurred; and (3) videos that make general accusations about TikTok unrelated to scams, such as accounts disqualification or followers not increasing.

Collect comments Leveraging the TikTok Researcher API, we also collected 1,256 comments associated with these 150 scam videos. Similar to Gilbert et al. [49], we collected comments for further sentiment analysis to determine how many support or disapprove of the video's content, as another check of the videos' validity. Of the 1,256 comments, we filtered out 93 that contained only emojis, leaving 1,163 comments for further analysis.

3.2 Data Analysis

Qualitative analysis of videos We thematically analyzed the collected videos both deductively and inductively. We made our codebook publicly available.[5] Since scams are a well-studied topic outside the TikTok context, we developed the initial version of our codebook deductively, to ground our analysis in existing conceptual frameworks and support comparability with the broader scam literature. This includes, for example, curating scam categories from Beals et al. [9], FTC [36], and various media articles [44, 46, 69]; drawing on the seven luring techniques in Stajano et al. [90]; and developing codes on anti-scam advice from FTC's resources [35].

We then expanded our codebook using inductive thematic analysis to capture insights specific to TikTok. E.g., we inductively developed codes for the scam sub-categories, scam's origin point, who the scammer was, and whether and to what extent the victim experienced any monetary harm (and if so how much). We also added new codes for anti-scam advice, and redressal mechanisms (e.g., 'product purchase and transaction-related advice' for advice and 'reporting to an entity' for redressal mechanisms).

Three researchers collaboratively analyzed the videos, meeting regularly to apply and refine the codebook, discuss ambiguous cases, and reach agreement on final code assignments. We used a consensus coding approach for the video content analysis, rather than relying on independent coders. Accordingly, we do not report inter-rater reliability for this stage, in line with HCI/CSCW qualitative coding guidelines [64]. We then finalized the codebook, grouping related codes into higher-level themes for scam types, lure techniques, scam identification, redressal, and advice.

Text analysis of video comments We conducted a large language model (LLM) assisted text analysis of 1,163 comments posted on the 150 videos to understand user reactions beyond engagement metrics. Specifically, we analyzed the sentiments conveyed through the comment texts to validate whether users engaging with the videos are generally supportive of the video's content (e.g., finding the video useful or informative, or relatable). We supplemented the analysis

[3]The complete list of information included hashtags, username, time when the video was created, video ID, comment count, comment text, like count, share count, view count, video description, and the voice-to-text conversion based on the audio.

[4]List of hashtags used: https://osf.io/yqxb9/files/wrpbu

[5]https://osf.io/yqxb9/files/wrpbu

of the supportive comments with thematic analysis using a consensus coding approach.

We leveraged the Gemini Flash 2.5 model[6] for the text analysis. Based on manual analysis of a few comment texts and video descriptions, we constructed a two-level classification task for the LLM: (1) Level-1 classification determines the overall stance of the comment, i.e., whether it is *supportive* of the video description; and (2) Level-2 classification further classifies supportive comments based on whether the user found the video *useful*, *relatable*, or *neither*. We carefully curated the LLM prompt to perform both classifications in a single call and to provide its reasoning, i.e., 'Chain-of-thought' prompting [109]. The prompt text includes class definitions, 'few-shot' examples, and rules for handling edge cases (see Table 3 for the full prompt).

To validate the LLM-assisted classification, we manually sampled 100 video-description-comment pairs. Two coders independently labeled them using the three level-1 class labels: *supportive*, *against*, and *neutral*. We achieved good inter-coder reliability (Cohen's Kappa) of 0.82 [43]. Similarly, for level-2 class labels for supportive comments: *relatable*, *useful*, and *neutral*, the inter-coder reliability (Cohen's Kappa) is 0.79. We discussed disagreements and resolved them to generate the final classification label for each comment. The curated LLM prompt achieves $\geq 94\%$ for precision, recall, and F1 scores on the level-1 classification task. Instances where the LLM's predictions differed from human labels were due to short, ambiguous comments, not errors.

3.3 Ethics Statement

We obtained approval and an exemption from full review from two Institutional Review Boards (IRBs). We also recognize that IRB approval does not guarantee ethical research, and there are various privacy and ethical risks associated with research using social media data [8], such as the difficulty of obtaining the data subject's explicit consent for their data to be used in research. To mitigate potential harms of exposing data to unintended audiences [41, 42], we stored only video URLs rather than downloading the videos to honor a data subject's withdrawal of consent: in case a video is deleted, it represents the data subject's wish to be removed from our dataset. We used the TikTok Researcher API to collect videos to follow the platform's policy and to mitigate potential privacy and ethical risks of using third-party scrapers. We also obscured users' faces in example images and paraphrased quoted content.

3.4 Limitations

We identify several potential limitations of our work. First, our analysis emphasizes depth over scale. Our goal was to unpack the nuances in user-reported scam videos on TikTok and the reactions they trigger, rather than quantify the prevalence of scams on TikTok. Second, our dataset may reflect a voluntary-response bias, as creators who actively post on TikTok may be more likely to share content about scams than more passive users. Nevertheless, our dataset still captures videos that users could easily find when seeking scam-related advice, which aligns with the study's objectives. Third, due to the API's restrictions, our dataset was limited to English-language content from US-based IP addresses. Consequently, our findings may not generalize to the diverse contexts prevalent in the global TikTok ecosystem. Future work could examine TikTok users' scam experiences across different cultural contexts.

Table 1: Mean (M) values of engagement metrics by the video type are reported, not in millions.

Video Type	*n*	M_{views}	M_{likes}	$M_{comments}$
Scam reporting	108	266	184	23
Anti-scam advice	15	55	57	11
Reporting and Advising	27	114	55	17

4 Findings

4.1 Video Characteristics

4.1.1 Video and Creator Metadata

Our final dataset consists of 150 videos where TikTok users report a scam experience and/or provide scam-related advice. Table 1 captures the engagement metrics of these videos, including the number of views, likes, and comments. On average, videos were 90 seconds long.The videos were posted by 129 unique user accounts (126 were public, three private). None of the accounts had a 'verified' account badge.[7] On average, these accounts received around 2.3M likes across their videos, maintained an average of 1,064 videos per profile, and had an average follower count of 69.6K. Table 4 shows the engagement metrics of the videos.

4.1.2 Comments Analysis

Among the 1,163 comments we analyzed using the LLM-assisted two-level classification technique, nearly 78% (902) were *supportive*, expressing agreement with the video creator or adding to the discussion. Of these, 43% (388/902) reported experiencing a similar scam or finding the video *relatable*, commonly recounting losses from past scams, warning about emerging trends, or sharing personal strategies against scams, e.g., *"I've noticed a few of those as well. TikTok also*

[6] https://deepmind.google/models/gemini/flash/ using VertexAI API; we also experimented with Gemini 3 Flash and 3 Pro models, but Gemini Flash 2.5 performed as well as Gemini 3 models on our task.

[7] A verified badge means that TikTok has confirmed the account belongs to the person or brand it represents [100]. The absence of a verified badge does not necessarily imply that an account is fake; it just means the account owner has not completed the verification process.

doesn't do much, so it's best to just block them." Another 11% (104/902) expressed that they found the videos useful or informative, some noting that it made them more aware of potential scams, e.g., *"Keep sharing. You are answering questions we didn't know we needed to know. Thank you."* The remaining supportive comments were generally positive or encouraging in nature, expressing concern about scam prevalence or offering help against scams, e.g., *"Ugh, that's so frustrating! I know it's discouraging, and it's not your fault. I'll see if I can report that page as well!"*

In the remaining 20% (233) of the comments, the users provided a neutral reaction or shared something irrelevant. Only 2% (26) of the comments contradicted or challenged the original video, e.g., "*If it was a good deal then what does it matter? It's just a bottle."* Overall, comment analysis shows high engagement with videos in the dataset. The predominantly positive sentiment expressed in the comments serves as a proxy for the video's legitimacy and trustworthiness.

4.2 Types of Scams Reported (RQ1)

Below, we summarize the six types of scams reported by TikTok users, ordered in frequency[8]: finance scams (54/33%), scams targeting influencers (49/30%), online shopping scams (36/22%), account-related scams (17/10%), romance scams (10/6%), and other miscellaneous scams (5/3%).

Table 2 in the Appendix further summarizes the definitions and examples of each scam type and maps them to similar, known scam types from prior work outside the TikTok context. While most of the scams reported by TikTok content creators can find parallels in prior work, our analysis shows how platform-specific affordances and influencer-centric culture can shape how scams are experienced and the harms they may cause. For instance, as shown by various influencer-targeted scams, the platform's influencer-centric culture has shifted the trust model. Instead of just impersonating distant institutions or authorities, scammers are now exploiting the parasocial bonds and community trust built between creators and followers, which complicates the traditional security advice of relying on verifying the source. Our analysis further surfaced a salient form of scam on TikTok, i.e., account-badge verification scam.

4.2.1 Scams involve financially-rewarding schemes

Most often, TikTok users reported finance scams where scammers try to lure people into financially rewarding schemes that turn out to be bogus. It includes four major types: mural scams, psychic reading scams, crowdfunding scams, and fake investment opportunities.

[8]The frequency only reflects how often we observe the scam type in our dataset and does not represent the estimated prevalence of these scam types on TikTok broadly. The scam types are not mutually exclusive, i.e., multiple types can be applied to a single video.

Mural scams are the most encountered finance scams in our dataset, mirroring the 'fake check scam' or 'fake buyer fraud' in prior work [9]. In mural scams, a scammer posing as an 'artist' sends a direct message, requesting to paint a mural of the victim from one of their profile pictures, as highlighted in Fig. 3 in the Appendix. They suggest a peculiar payment arrangement in which a 'client' would send an e-check to the victim for the full amount, after which the victim forwards half to the scammer for painting materials. However, the e-check later remains pending or bounces, leaving the victim with no money.

Psychic reading scams are similar to the fortune-telling fraud reported in prior work [9], where a scammer claims to tell the victim's fortune in return for a fee. Sometimes psychic reading scams overlap with impersonation scams. For example, T18 reported contacting a scammer via TikTok direct messages (see Fig. 5 in Appendix), believing she was communicating with a legitimate influencer whose content she liked. The scammer quickly requested payment for a psychic reading, raising her suspicion. After T18 stopped responding, the scammer started liking and commenting on all her videos to get her attention. Upon further investigation, she discovered that the scammer account's videos were copied from another TikTok page.

Crowdfunding scams, also known as the 'charitable solicitations' from prior work [9], often involve crowdfunding for bogus personal causes such as fake illness or sick pets. Moreover, TikTok users report *fake investment opportunities*, similar to miscellaneous investments, Ponzi schemes, and crypto scams mentioned in prior work [9, 70]. TikTok users reported that scammers try to lure them into easy-money investment schemes. For instance, T22 shared their encounter with a fraudulent brokerage account that promised $20K in virtual currency upon activation, but required a $300 upfront payment. They contrasted it with legitimate brokerage companies like Robinhood, which request minimal verification fees like $2 to activate an account.

Across reported financial scams, scammers primarily initiated contact via TikTok direct messages, with a few instances of mural scams reported on Instagram direct messages. Monetary losses were limited: only 3 of 54 users reported losses, with the largest at approximately $35 in a crowdfunding scam.

4.2.2 Scams targeting influencers and influencer culture

Uniquely tied to TikTok's influencer-centric culture, TikTok users, especially those who are active content creators, reported scams that target their own celebrity and fan bases. The specific schemes include account impersonations, fake collaborations, and false promises of growing the account by a social media brand manager.

TikTok users reported *impersonation scams* in multiple ways. Most commonly, creators shared screenshots of impersonated accounts from scammers and warned their followers about it. TikTok users also identified characteristics of impersonation scams, e.g., scammers would often use the original creator's profile picture and the same name with slight variation, as highlighted in Figure 5(b) in the Appendix. T31 shared, *"My account gets cloned all the time by scammers. I don't slide into anyone's DM to sell or promote anything. If someone messaged you from any profile that looks like mine, it's not me"* (see Figure 6(a)) in the Appendix. In other cases, the TikTok user encountered a fake profile of a creator they like; when they followed those fake accounts, the scammers asked for money.

In *fake collaboration scams*, the influencer reported being approached by companies to endorse their products, but was required to purchase them first; in other cases, the influencer was never compensated for sales generated by their endorsement. Two creators, T97 and T145, reported the same vendor who approached them via email under a 'UGC content creation request' (see Figure 6(b)). Fake collaboration scams share some similarity with 'business opportunities fraud ' [9] and the employment/job fraud [36, 70], where victims are promised support to build a business or gain employment, but companies fail to fulfill that promise.

Other TikTok users reported *social media brand manager scams*, in which the scammer pretended to be a social media brand management agency and promised to increase the creator's followers and view counts, and to help with brand collaborations. T61 reported that a scammer contacted him, offering event opportunities and technical support to grow his account. However, when he checked the scammer's TikTok page, the account was banned, indicating that the scammer had violated TikTok's community guidelines. This type of scam parallels the 'business coaching scam' [9], 'sports agent and endorsement scams' [36], and 'fake LinkedIn recruiters' [70], where companies promise business training and tools but fail to deliver, frequently charging high fees.

Across influencer-targeting scams, the scammer tends to be a business or organization that reached out through TikTok direct messages or email. Users reported the second-most financial loss (only after online shopping scams), with maximum loss being $1000 in a fake collaboration scam.

4.2.3 Online shopping scams involving fake products

Online shopping scams ranked third among the most commonly reported scams. TikTok users reported fraudulent experiences when buying a product or service through 'TikTok Shop'[9] or through advertisements on their 'For You Page.' Users reported receiving *fake or poor quality products*, similar to the 'worthless products' scam in prior work [9], which involves false or misleading information about the products or goods that exaggerate the claims or turn out to be worth much less than anticipated. For instance, T10 reported being *"duped for a fake Stanley Cup."* Other scams in this category involved *unauthorized billing* for products or services and *undelivered goods*. T62 reported receiving unauthorized billing for a product that had not been delivered. She was charged an extra $50 beyond her original order of a similar amount. Although she received emails confirming the order, sharing details about tracking the package, and notifying her of a delay, she never received the product.

All in all, users reported the most financial loss to online shopping scams, with the maximum amount being $3000. Unlike the individual creators behind finance-related scams, scammers in this category were mainly businesses.

4.2.4 Account-related scams fueled by badge verification

Less often, TikTok users reported account-related scams in which they were tricked into giving away personal information and sometimes lost money, either through fake *account badge verification* offers or *account hijacking*.

On TikTok, a verified badge is a blue check mark that appears next to a user's account name, indicating that the account is authentic. TikTok grants these badges to accounts that meet specific criteria, including active usage and recognition in multiple news sources [100]. In *account verification badge* scams, users reported receiving emails impersonating TikTok that asked them to verify their accounts to receive the verification badge. The email had a legitimate-looking subject line (e.g., 'RE: Badge Application - #831947') from 'TikTok Support,' a TikTok logo, and a 'Verify Now' button in red. In some cases, the scam led to *account hijacking*. For example, T77 reported that she clicked the 'verify now' button and entered credentials on a fake TikTok login page, resulting in unauthorized access to her account and subsequent takeover. This scam represents a TikTok-salient variant that combines institutional impersonation [9] and phishing [36] to exploit creator verification incentives.

Another account-related scam involves *tagging* users in fake CashApp or lottery schemes. TikTok users reported receiving direct messages such as *"you have been selected as the winner"* for some money contests, mirroring the 'prizes, sweepstakes, and lotteries scams' in prior work [36]. To appear legitimate, scammers often used the CashApp logo, a popular digital wallet (see Figure 7(b) in the Appendix).

Across account-related scams, scammers mostly operated as individual creators rather than businesses. They reached victims via email, TikTok direct messages, or curated content on the 'For You Page.' We did not observe any financial losses.

4.2.5 Sugar daddy scams: romance scams tailored to TikTok demographics

Though less common, romance scams, a well-known scam type, were also reported by TikTok users in both *traditional*

[9]'TikTok Shop' is an in-app personalized commerce solution where brands and creators sell products directly on TikTok, through a suite of in-app shopping touchpoints such as a shop page and live shopping [98].

forms and as *sugar daddy scams* tailored to TikTok's user base. As in prior work [9, 36, 113, 114], conversations often began with a seemingly innocent direct message, followed by love bombing and eventually financial requests. For *sugar daddy scams*, demographics are the distinctive factor: scammers posed as wealthy, older men who initiated contact via direct messages, offering money for companionship or a romantic relationship. Unlike the consensual sugar daddy relationships [79], these interactions resulted in requests for money, mirroring romance scam patterns. TikTok's user base, which skews toward young women in the U.S. [92], may contribute to the rise of such scams on TikTok.

The romance scams reported by TikTok users are also not unique to TikTok. For example, T64 reported that a scammer reached out to her via direct message on Instagram, asking her to hang out and then asked to move the conversation to another platform, such as Google Chat or text message. The similar pattern—in which initial contact occurred through direct messaging on social media and was followed by requests to move to another platform—was reported by other TikTok users and mirrors prior work that documents the trajectory of sugar daddy [45, 59, 80] and pig-butchering scams [71].

We also observed miscellaneous scams, including *AI-driven deepfakes*, *health-related scams*, and *educational financial support scams*. T98 shared an example of *AI-driven deepfake* scam (see Fig. 4 in Appendix), in which scammers manipulated her original video to falsely promote a fraudulent product. T98 expressed concern that such altered content could damage her reputation, as followers might believe she endorsed products that never arrive or are of poor quality.

4.3 How User-Reported Scams Manifest (RQ2)

Going beyond the types of scams, TikTok users also shared where the scams came from and their analysis of how scammers tried to defraud them. In this section, we summarize findings on these scam origin points and tactics used by scammers, mapped to the seven principles in Stajano et al. [90].

4.3.1 Scams commonly originate from TikTok direct messages

We found that most reported scams originated from TikTok direct messages (30%) or the TikTok 'For You Page' (30%). Users also reported encountering the scam through TikTok shops (13%), video comments (2%), post tagging (2%), and TikTok Live feeds (2%). External origin points include phishing emails (5%), direct messaging on other social media platforms (3%), and phone calls or text messages (2%).

Mapping these origin points to scam types (see Figure 1), we observed that TikTok direct messages are a common origin point for account-related, finance, influencer-targeting, and romance scams. Sometimes, influencer-targeting and account-related scams also occur through email. For online shopping

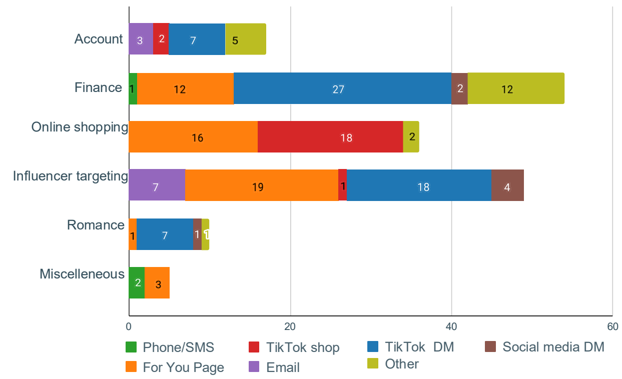


Figure 1: The distribution of origin points for scam types.

scams, 'For You Page' (FYP) and 'TikTok Shop' are the most common origin points.

4.3.2 Luring Techniques Used By Scammers

Some TikTok users highlighted the psychological techniques scammers use, which we analyzed using the seven principles by Stajano and Wilson [90]. We found that the two techniques, 'need and greed' and 'distraction,' often occur together.

The 'need and greed' principle is most often reflected in reported experiences of finance, influencer-targeting, and romance scams. Under this principle, scammers exploit individuals' needs and desires, such as financial need or emotional vulnerability. For example, in psychic reading scams, users described being offered fortune-telling services in exchange for payment, indicating scammers' exploitation of people's emotional need to know about their future. In influencer-targeting scams, scammers exploit creators' desire to grow their accounts by making false promises of increasing followers and presenting fake collaboration opportunities.

Users also described experiences that reflect scammers' exploitation of the 'distraction' principle, in which scammers divert the victim's attention and carry out their deceptive act in the background. The 'distraction' principle is particularly salient in videos reporting online shopping scams. Users recounted being drawn to low-cost products, only for scammers to sneakily add shipping costs or dupe them with counterfeit products. For instance, in T62's case (§4.2.3), she reported being lured by an ad for what appeared to be a high-quality makeup product on a polished website. After placing an order for approximately \$50, the company automatically charged her for an additional \$50 product without consent, and neither product was ever delivered.

Scammers also leverage 'need and greed' and 'distraction' together to increase the effectiveness of their lures. This is the case in reported fake collaboration scams, for example, where creators described receiving emails impersonating TikTok representatives (distraction) that offered verified account badges as an incentive (need and greed). T77 reported falling victim to such a scam (§4.2.4).

4.4 Scam Identification and Redress (RQ3)

4.4.1 How TikTok Users Identify Scams

Here, we summarize the signals TikTok users leverage to identify a scam. Among the four signals documented by the FTC [35], we found that TikTok users mostly relied on scammers' requests to pay money in specific ways as a signal. For instance, users reported that in mural scams, when scammers request that the victim deposit an e-check from a third-party app, it is a red flag. Others shared instances of scammers' impersonation attempt (e.g., in account badge verification scams) or when scammers pressure them to act immediately.

Beyond these well-known signs, our analysis surfaces additional signals that might be more specific to TikTok's community culture and norms. Users reported identifying scams based on certain communication patterns, e.g., a generic template used in mural scams when scammers reached out via direct messaging (DM): *"I was thinking about using one of your photos for a fine art project I'm working on for my client, and you'll also get paid for it as a bonus."* T73 helped their followers understand their own communication patterns in contrast to the scammer: *"There is a fraudulent profile posing as me (creator) on TikTok, and reaching out to people offering a special deal. Please note I will NEVER conduct bookings via DM on any socials, and I will not call you beloved."* Users also highlighted linguistic red flags, such as grammatical errors in brand names or unconventional language. For example, T39 reportedly identified a scammer by observing that *"(when texting) big paragraphs have perfect English, but smaller ones are full of grammar errors."* T43 shared that the *"scammer used unconventional phrases and words like 'dear', 'sweetie', and 'bucks' - Americans don't use those words."*

Account discrepancy is another sign that TikTok users use to detect scams. A lack of social media or internet presence could trigger suspicion. For instance, T70 shared that when a scammer reached out in a mural scam, they checked the scammer's profile and found that it *"had three photos and only 15 followers on Instagram."* Other TikTok users detected scams by spotting mismatched email domain names, differences in the number of followers between the actual user's and the scammer's profiles, or profile details that did not match in cases of impersonation.

Product and pricing misrepresentation is another sign users leverage to identify a scam, particularly online shopping scams. E.g., T30 reported that when she received a product, she found that a product advertised for eczema treatment was not actually recommended for that use, as revealed in the fine print. Other users paid close attention to pricing and became suspicious when prices seemed unusually high or unrealistically low. They compared prices elsewhere on the internet and often found similar products at lower prices, helping them avoid buying from scammers.

4.4.2 How TikTok Users Redress Scams

Our analysis shows that TikTok users employed five major redressal strategies against scams: public disclosure, reporting to entities, in-platform protective measures, information seeking, and community mobilization.

By sharing scam-related videos, TikTok serves as a platform for public disclosure and awareness-raising about scams. The user-reported scam videos often included evidentiary content, such as screenshots of interactions with scammers or direct exposure of their identities, along with advice on how to mitigate such scams. The advice falls into three broad categories: a) product purchase and transaction-related advice (e.g., researching products and sellers thoroughly before making a purchase), primarily for online shopping scams; b) scammer profile and communication style (e.g., highlighting linguistic red flags such as unusual phrasing), primarily for finance and influencer-targeting scams; c) generic security advice, such as avoiding unfamiliar links.

When losses occur, users reach out to various entities to seek redress, including TikTok, payment vendors such as PayPal, and, in some cases, the scammers themselves. Those who contacted TikTok often reported experiencing a lack of support and no visible action taken against the scammers. For example, T23 reported multiple TikTok accounts impersonating her profile, but TikTok concluded there was no violation of its community guidelines and simply suggested she block those accounts. Similarly, users who reached out directly to scammers rarely, if ever, received a response. In contrast, contacting payment vendors such as PayPal or credit card companies was often far more effective, with many users successfully securing refunds through these channels.

In addition to seeking institutional redress, users applied various in-platform protective measures, most commonly blocking scammers. For instance, T45 mentioned blocking an account that reached out to paint their mural. Others created new accounts or used additional security measures like Face ID to log into the account.

Users also engaged in information-seeking behaviors, such as consulting Reddit forums to learn about scams or using Google's reverse image search to verify a scammer's identity. T43 shared that they used Google's reverse image search to investigate a scammer's profile picture and discovered that it had been sourced from another website. Community support on TikTok emerged as another prominent strategy, particularly for impersonation scams. Creators encouraged their followers to report fraudulent profiles, while followers alerted creators to accounts impersonating them.

We map the various redressal mechanisms to the scam types in Figure 2. While public disclosure was common across all scam types (and so not shown), other redressal strategies vary.

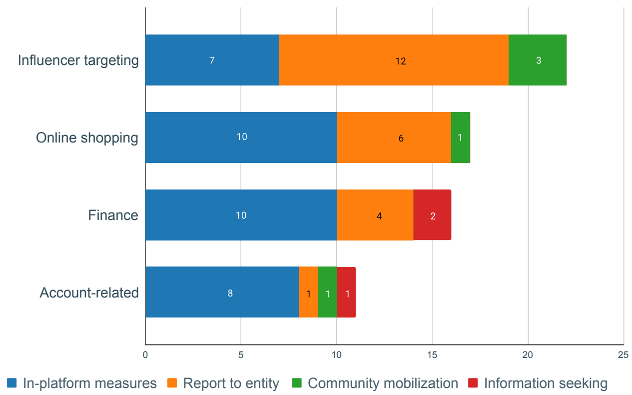


Figure 2: Distribution of redressal mechanisms across prominent scam categories.

Account-related scams prompted multifaceted responses, including in-platform protections, reporting to relevant entities, and community mobilization. Online shopping and finance scams more often led to efforts to recover funds through payment providers and in-platform protections. Influencer-targeted scams saw higher reporting levels to platforms, especially TikTok. Additionally, information-seeking behaviors were notably high in finance scams, where users questioned the legitimacy of messages or offers.

5 Discussion

5.1 Key Insights

Comparing scams reported by TikTok users with prior work Scams are a longstanding societal problem that predates the Internet. Most of the scam types in our dataset §4.2 parallel categories from Beals et al.'s framework [9] and FTC fraud types [36], along with prior work on specific scam types such as online shopping [54, 66], romance [105, 113, 114], pig-butchering [71], deepfake-enabled scams [118], and cryptocurrency scams [117] (see Table 2). Nonetheless, these scam types manifest differently on TikTok, shaped by the platform's affordances and influencer-centric culture, a defining characteristic of SVPs [6, 82]. For instance, while impersonation scams mirror impostor scams [38], on TikTok these scams specifically exploit the influencer–follower trust by creating fake profiles of those influencers and requesting money from followers (§4.2.2). Account badge verification scams are particularly salient on TikTok: although they combine elements of institutional impersonation [9] and phishing [36], they target influencers through phishing emails that impersonate TikTok and offer verification badges, ultimately locking creators out of their accounts after a password reset (§4.2.4). Building on recent work highlighting the digital safety issues faced by content creators [84, 94], we show that being targeted by scammers further increases influencers' risk exposure.

How TikTok scams occur: same underlying principles, platform-specific entry points We observe that scammers' tactics in our dataset generally align with the seven luring principles outlined by Stajano and Wilson [90], though our findings highlight that their application varies across scam types. For instance, the 'need and greed' principle—exploiting victims' desires—is pervasive across scam types, whereas the 'distraction' principle—targeting victims when they are preoccupied by something else—appears predominantly in online shopping scams. Scammers may simultaneously employ multiple lure principles, such as combining 'need and greed' with 'distraction' in account badge verification scams (§4.3.2).

Rather than introducing new psychological strategies, TikTok scams adapt established lures to the platform affordances. Most scams reported by TikTok users originated within the platform, commonly via direct messages or the For You Page (§4.3.1). This builds on prior research on '#DMMe" scams on Instagram [101], where scammers prompt users to initiate direct messages, which are then exploited to extract money or sensitive information under the guise of offering legitimate services. The presence of scam-related content on the FYP may also complicate users' understanding of how the algorithm curates content. Such encounters can introduce uncertainty into users' algorithmic folk theories [58] and potentially contribute to acts of algorithmic resistance [102].

Scam security folklore and anti-scam advice are rooted in personal experience TikTok users' reported experiences with scams, ways to identify them, redressal mechanisms, and advice on how to stay safe from scams together form a type of security folklore. Some recommendations—such as caution with unfamiliar links (§4.4.2)—echo established security advice in prior work [18, 55, 81]. Others seek to promote healthy shopping habits to mitigate online shopping scams or highlight scammers' communication patterns and profile characteristics. The communication patterns, recognized by TikTok users, align with FTC's guidelines on how to avoid a scam [35], while also being tailored to the TikTok audience. For instance, creators share specific scam templates (e.g., those used for mural scams) or highlight inconsistencies in scammer profiles by comparing them with their own accounts (§4.4.1). This community-driven advice and redress could complement TikTok's existing anti-scam efforts, e.g., the Safety Center [96], which lists known scams but could also incorporate recurring communication and behavioral patterns identified in this work that may apply to emerging scams.

While the provision of anti-scam advice can take many forms, we observe that across all videos in our dataset that share anti-scam advice, a majority also report on a specific scam (Table 1), often grounded in personal experience. Creators give detailed accounts of the specific fake products they fell for to contextualize their advice on researching the sellers (§4.2.3), or draw from their interactions with fraudulent brokerage services when calling out specific communication patterns used by scammers (§4.4.2). The nature of TikTok anti-scam advice, often driven by personal stories, also reflects the effectiveness of storytelling about negative incidents [78] and folklore [34], as laypeople rely on informal lessons more than official sources for learning about security.

TikTok as an avenue for anti-scam peer support and studying pro-security advice Our research builds upon a growing body of scholarship that uses social media data—such as those from Facebook [72], Reddit [15, 23], and YouTube [16, 17, 68]—to study particular types of scams, such as cryptocurrency scams [23], sextortion scams [104], as well as other trust and safety issues, such as image-based sexual abuse [110]. A coherent theme across this literature is how community members provide peer support by sharing their vulnerabilities and providing actionable guidance to help others prevent or recover from similar forms of victimization [23, 110]—a pattern we also observe in our findings. Prior work by Wei et al. used TikTok as a source for studying anti-security and anti-privacy advice [111]. In contrast, our work, which focused on pro-security advice, observed that TikTok serves as both a site where scams may originate and a space where victims can support one another. This duality presents nuanced dynamics for victims to leverage the platform for peer support while also confronting the risk of (re)traumatization associated with their experiences.

5.2 Recommendations

Systemizing platform security support TikTok's "Safety Checkup" feature aims to prevent account verification scams by allowing users to link and verify contact information, enable two-step verification, manage trusted devices, and add a passkey. However, these security features are optional, despite their potential to strengthen account security following an account verification scam. Similarly, 2-step verification is optional for other SVPs such as Instagram [2]. Furthermore, TikTok relies on social authentication for account recovery, involving friends in the verification process, unlike Instagram's video verification, which introduces additional usability overhead. Given the prevalence of account-related scams in our dataset, requiring completion of key Safety Checkup steps could reduce their impact. SVPs and OS providers could also collaboratively develop on-device support and offer timely, privacy-preserving scam detection and notification across applications. For instance, they can leverage Android's on-device AI, already used for detecting phone scams [50], and privacy-preserving model update techniques [119].

Usability of automated content filtering. We found that more scams originate from direct messages than from any other vector. On TikTok, most users access direct messages through either their "Inbox" or "Message Requests." However, safety features such as spam filtering are only available to users with over 10,000 followers who have access to the enhanced creator inbox. As a result, distinguishing genuine from fraudulent messages can be more difficult for most users who do not meet this follower threshold. One promising feature is TikTok's content labeling, which marks videos as AI-generated, as containing unverified or false information, or as originating from state media—either automatically or manually by the creator. Prior work shows that such labels can help users assess information credibility [32]. However, without labels, users may mistakenly believe content is authentic [74]. Automated content filters could also be better personalized to meet community needs, a task creators often handle manually, as noted on YouTube's creator safety page [1]. Extending labeling and filtering efforts to encompass scams and anti-scam advice may help users better protect themselves on SVPs.

Leveraging influencers, multi-channel network (MCN) brand managers, and storytelling in scam experiences We found that scammers often exploit followers' trust in influencers and creators' visibility on TikTok. Conversely, this suggests significant potential to leverage influencers as key opinion leaders to raise awareness of scams and promote protective behaviors among users. Prior work also highlights that peer storytelling is one of the most effective ways to generate awareness and educate people by sharing of personal experiences with scams [78]. Implementing this strategy requires TikTok to strengthen its safety mechanisms against emerging threats, such as AI-generated deepfakes, while better protecting creators. One promising avenue is the platform's existing multi-channel network (MCN) ecosystem, where TikTok encourages brand managers to acquire, educate, and incubate creators within the TikTok Shop [99]. Although we found cases of scammers impersonating MCN brand managers to defraud creators, this risk highlights the need for stronger MCN safeguards. Verified MCNs could help creators identify legitimate collaborations while providing scam-awareness training, strengthening the creator–follower ecosystem.

6 Conclusion

We analyzed 150 TikTok videos of users' reported experiences with fraudulent content and scams. Our study reveals unique scam types targeting SVPs, tactics scammers use, signals to identify scams, and advice to mitigate them. While the scam categories and luring techniques are not entirely new, we show how TikTok's influencer culture and platform features amplify these scams. Additionally, TikTok serves as a venue for anti-scam advice and peer support. Based on our findings, we provide recommendations for systematizing platform support for scam prevention and utilizing influencer culture to increase scam awareness. Our work has implications for future research on fraudulent content and on effective strategies for identifying and mitigating scams.

Acknowledgments

The research is partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2092 CASA - 390781972 and a Google Academic Research Award on Trust and Safety (2024; ID 00029925).

References

[1] Creator Safety Center: Support & Resources - YouTube Creators — youtube.com. https://www.youtube.com/creators/safety/#growing-your-channel-safely. [Accessed 11-09-2025].

[2] Instagram Account Safety & Security | Secure Your Account | About Instagram — about.instagram.com. https://about.instagram.com/safety/account-security. [Accessed 11-09-2025].

[3] Crystal Abidin. Visibility Labour: Engaging with Influencers' Fashion Brands and #OOTD Advertorial Campaigns on Instagram. *Media International Australia*, 2016.

[4] Bhupendra Acharya and Thorsten Holz. An Explorative Study of Pig Butchering Scams. *arXiv preprint arXiv:2412.15423*, 2024.

[5] Bhupendra Acharya, Muhammad Saad, Antonio Emanuele Cinà, Lea Schönherr, Hoang Dai Nguyen, Adam Oest, Phani Vadrevu, and Thorsten Holz. Conning the Crypto Conman: End-To-End Analysis of Cryptocurrency-Based Technical Support Scams. In *Proc. IEEE S&P*, 2024.

[6] Juan Miguel Alcántara-Pilar, María Eugenia Rodriguez-López, Zoran Kalinić, and Francisco Liébana-Cabanillas. From Likes to Loyalty: Exploring the Impact of Influencer Credibility on Purchase Intentions in TikTok. *Journal of Retailing and Consumer Services*, 2024.

[7] Holly Avella. "TikTok≠ Therapy": Mediating Mental Health and Algorithmic Mood Disorders. *New Media & Society*, 2024.

[8] Kyle Beadle, Kieron Ivy Turk, Aliai Eusebi, Mindy Tran, Marilyne Ordekian, Enrico Mariconti, Yixin Zou, and Marie Vasek. SoK: A Privacy Framework for Security Research Using Social Media Data. In *Proc. IEEE S&P*, 2025.

[9] Michaela Beals, Marguerite DeLiema, and Martha Deevy. Framework for a Taxonomy of Fraud. *Financial Fraud Research Center*, 2015.

[10] Mario Beluri, Bhupendra Acharya, Soheil Khodayari, Giada Stivala, Giancarlo Pellegrino, and Thorsten Holz. Exploration of the Dynamics of Buy and Sale of Social Media Accounts. In *Proc. IMC*, 2025.

[11] Sruti Bhagavatula, Lujo Bauer, and Apu Kapadia. "Adulthood Is Trying Each of the Same Six Passwords that You Use for Everything": The Scarcity and Ambiguity of Security Advice on Social Media. *Proc. ACM Hum.-Comput. Interact.*, 2022.

[12] Aparajita Bhandari and Sara Bimo. Why's Everyone on TikTok Now? the Algorithmized Self and the Future of Self-Making on Social Media. *Social media+ society*, 2022.

[13] Trevor Boffone. Introduction: The Rise of TikTok in US Culture. In *TikTok Cultures in the United States*. Routledge, 2022.

[14] Marcus Bösch and Tom Divon. The Sound of Disinformation: TikTok, Computational Propaganda, and the Invasion of Ukraine. *New Media & Society*, 2024.

[15] Elijah Bouma-Sims, Hiba Hassan, Alexandra Nisenoff, Lorrie Faith Cranor, and Nicolas Christin. "It Was Honestly Just Gambling": Investigating the Experiences of Teenage Cryptocurrency Users on Reddit. In *Proc. SOUPS*, 2024.

[16] Elijah Bouma-Sims, Lily Klucinec, Mandy Lanyon, Julie Downs, and Lorrie Faith Cranor. The Kids Are All Right: Investigating the Susceptibility of Teens and Adults to YouTube Giveaway Scams. In *NDSS*, 2025.

[17] Elijah Bouma-Sims and Brad Reaves. A First Look at Scams on Youtube. In *Proc. MADWeb*, 2021.

[18] Karoline Busse, Julia Schäfer, and Matthew Smith. Replication: No One Can Hack My Mind Revisiting a Study on Expert and Non-Expert Security Practices and Advice. In *Proc. SOUPS*, 2019.

[19] Mark Button, Carol McNaughton Nicholls, Jane Kerr, and Rachael Owen. Online Frauds: Learning from Victims Why They Fall for These Scams. *Australian & New Zealand Journal of Criminology*, 2014.

[20] M. Ariel Cascio, Eunlye Lee, Nicole Vaudrin, and Darcy A. Freedman. A Team-Based Approach to Open Coding: Considerations for Creating Intercoder Consensus. *Field Methods*, 2019.

[21] Laura Cervi and Tom Divon. Playful Activism: Memetic Performances of Palestinian Resistance in TikTok #Challenges. *Social Media+ Society*, 2023.

[22] Hongliang Chen, Christopher E. Beaudoin, and Traci Hong. Securing Online Privacy: An Empirical Test on Internet Scam Victimization, Online Privacy Concerns, and Privacy Protection Behaviors. *Computers in Human Behavior*, 2017.

[23] Andrew Childs. 'I Guess That's the Price of Decentralisation…': Understanding Scam Victimisation Experiences in an Online Cryptocurrency Community. *International Review of Victimology*, 2024.

[24] Sauvik Das, Adam DI Kramer, Laura A Dabbish, and Jason I Hong. The Role of Social Influence in Security Feature Adoption. In *Proc. CSCW*, 2015.

[25] Daniel Delmonaco, Samuel Mayworm, Hibby Thach, Josh Guberman, Aurelia Augusta, and Oliver L Haimson. "What Are You Doing, TikTok?": How Marginalized Social Media Users Perceive, Theorize, and "Prove" Shadowbanning. *Proc. ACM Hum.-Comput. Interact.*, 2024.

[26] Michael Ann DeVito. How Transfeminine TikTok Creators Navigate the Algorithmic Trap of Visibility via Folk Theorization. *Proc. ACM Hum.-Comput. Interact.*, 2022.

[27] Stacy Jo Dixon. Most popular social networks worldwide as of february 2025, by number of monthly active users. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/, 2025. Accessed: 2025-04-15.

[28] Brooke Erin Duffy. *(Not) Getting Paid to Do What You Love: Gender, Social Media, and Aspirational Work*. Yale University Press, 2017.

[29] Alan Dundes. Folk Ideas as Units of Worldview. *The Journal of American Folklore*, 1971.

[30] Paul Dunphy, Vasilis Vlachokyriakos, Anja Thieme, James Nicholson, John McCarthy, and Patrick Olivier. Social Media as a Resource for Understanding Security Experiences: A Qualitative Analysis of #Password Tweets. In *Proc. SOUPS*, 2015.

[31] Ran Elgedawy, John Sadik, Anuj Gautam, Trinity Bissahoyo, Christopher Childress, Jacob Leonard, Clay Shubert, and Scott Ruoti. Security Advice for Parents and Children About Content Filtering and Circumvention as Found on YouTube and TikTok. In *Proc. USEC*, 2025.

[32] Ziv Epstein, Antonio Alonso Arechar, and David Rand. What Label Should Be Applied to Content Produced by Generative AI? *PsyArXiv*, 2023.

[33] Michael Fagan and Mohammad Maifi Hasan Khan. Why Do They Do What They Do?: A Study of What Motivates Users to (Not) Follow Computer Security Advice. In *Proc. SOUPS*, 2016.

[34] Matthias Fassl, Alexander Ponticello, Adrian Dabrowski, and Katharina Krombholz. Investigating Security Folklore: A Case Study on the Tor over Vpn Phenomenon. *Proc. ACM Hum.-Comput. Interact.*, 2023.

[35] Federal Trade Commission. How to avoid a scam. https://consumer.ftc.gov/articles/how-avoid-scam, 2022. Accessed: 2025-04-15.

[36] Federal Trade Commission. Consumer sentinel network: Product service code full descriptions. https://www.ftc.gov/system/files/ftc_gov/pdf/CSNPSCFullDescriptions.pdf, 2023. Accessed: 2025-02-04.

[37] Federal Trade Commission. Help the FTC fight impersonation scams. https://consumer.ftc.gov/consumer-alerts/2025/04/help-ftc-fight-impersonation-scams?utm_source=govdelivery, 2025. Accessed: 2025-04-04.

[38] Federal Trade Commission. New FTC data show a big jump in reported losses to fraud to $12.5 billion in 2024. https://www.ftc.gov/news-events/news/press-releases/2025/03/new-ftc-data-show-big-jump-reported-losses-fraud-125-billion-2024, 2025. Accessed: 2025-04-15.

[39] Federal Trade Commission. Top scams of 2024: investment scams top the list, followed by imposter scams. https://consumer.ftc.gov/consumer-alerts/2025/03/top-scams-2024, March 2025. Accessed: 2025-08-24.

[40] Ana Ferreira, Lynne Coventry, and Gabriele Lenzini. Principles of Persuasion in Social Engineering and Their Use in Phishing. In *Proc. HAS*, 2015.

[41] Casey Fiesler and Nicholas Proferes. "Participant" Perceptions of Twitter Research Ethics. *Social Media+ Society*, 2018.

[42] Casey Fiesler, Michael Zimmer, Nicholas Proferes, Sarah Gilbert, and Naiyan Jones. Remember the Human: A Systematic Review of Ethical Considerations in Reddit Research. *Proc. ACM Hum.Comput. Interact.*, 2024.

[43] Joseph L Fleiss, Bruce Levin, and Myunghee Cho Paik. *Statistical Methods for Rates and Proportions*. John Wiley & Sons, 2013.

[44] Amelia Fletcher, Gregory S Crawford, Jacques Crémer, David Dinielli, Paul Heidhues, Michael Luca, Tobias Salz, Monika Schnitzer, Fiona M Scott Morton, Katja Seim, et al. Consumer Protection for Online Markets and Large Digital Platforms. *Yale J. on Reg.*, 2023.

[45] Alyssa Forato. Are the sugar daddies in your DMs real? A cybersecurity expert weighs in. https://fashionjournal.com.au/life/sugar-daddy-instagram-dms/, January 2025. Accessed: 2025-06-04.

[46] Janet Fowler. 10 common scams targeted at teens. https://www.investopedia.com/financial-edge/1012/common-scams-targeted-at-teens.aspx, November 2022. Accessed: 2025-06-02.

[47] FTC. FTC issues orders to social media and video streaming platforms regarding efforts to address surge in advertising for fraudulent products and scams. https://www.ftc.gov/news-events/news/press-releases/2023/03/ftc-issues-orders-social-media-video-streaming-platforms-regarding-efforts-address-surge-advertising, 2023.

[48] Naomi Genota. 10 proven ways to spot influencer scams & brands that prey on small influencers. https://naohmsmedia.com/influencer-scams/, 2024. [Accessed 11-09-2025].

[49] Meira Gilbert, Miranda Wei, and Lindah Kotut. "TikTok, Do Your Thing": User Reactions to Social Surveillance in the Public Sphere. In *Proc. SOUPS*, 2025.

[50] Google. Use Scam Detection - Phone app Help — support.google.com. https://support.google.com/phoneapp/answer/15654065?hl=en. [Accessed 05-06-2025].

[51] Paul Harrigan, Timothy M Daly, Kristof Coussement, Julie A Lee, Geoffrey N Soutar, and Uwana Evers. Identifying Influencers on Social Media. *International Journal of Information Management*, 2021.

[52] Camille Harris, Amber Gayle Johnson, Sadie Palmer, Diyi Yang, and Amy Bruckman. " Honestly, I Think TikTok Has a Vendetta Against Black Creators": Understanding Black Content Creator Experiences on TikTok. *Proc. ACM Hum.-Comput. Interact.*, 2023.

[53] Samantha Hautea, Perry Parks, Bruno Takahashi, and Jing Zeng. Showing They Care (or Don't): Affective Publics and Ambivalent Climate Activism on TikTok. *Social Media+ Society*, 2021.

[54] Sherry He, Brett Hollenbeck, Gijs Overgoor, Davide Proserpio, and Ali Tosyali. Detecting Fake-Review Buyers Using Network Structure: Direct Evidence from Amazon. *Proc. National Academy of Sciences*, 2022.

[55] Iulia Ion, Rob Reeder, and Sunny Consolvo. "No One Can Hack My Mind": Comparing Expert and Non-Expert Security Practices. In *Proc. SOUPS*, 2015.

[56] Eleonora Isaia, Noemi Oggero, and Davide Sandretto. Is Financial Literacy a Protection Tool from Online Fraud in the Digital Era? *Journal of Behavioral and Experimental Finance*, 2024.

[57] Leanne Johnstone and Cecilia Lindh. Sustainably Sustaining (Online) Fashion Consumption: Using Influencers to Promote Sustainable (Un)planned Behaviour in Europe's Millennials. *Journal of Retailing and Consumer Services*, 2022.

[58] Nadia Karizat, Dan Delmonaco, Motahhare Eslami, and Nazanin Andalibi. Algorithmic Folk Theories and Identity: How TikTok Users Co-Produce Knowledge of Identity and Engage in Algorithmic Resistance. *Proc. ACM Hum.-Comput. Interact.*, 2021.

[59] Maeve Kerr-Crowley. What it's really like to have a sugar daddy. https://fashionjournal.com.au/life/what-its-like-to-have-sugar-daddy/, April 2022. Accessed: 2025-06-04.

[60] Jeehyun Jenny Lee and Jin Lee. # StopAsianHate on TikTok: Asian/American Women's Space-Making for Spearheading Counter-Narratives and Forming an Ad Hoc Asian Community. *Social Media+ Society*, page 20563051231157598, 2023.

[61] Xigao Li, Anurag Yepuri, and Nick Nikiforakis. Double and Nothing: Understanding and Detecting Cryptocurrency Giveaway Scams. In *NDSS*, 2023.

[62] Chen Lou and Shupei Yuan. Influencer Marketing: How Message Value and Credibility Affect Consumer Trust of Branded Content on Social Media. *Journal of Interactive Advertising*, 2019.

[63] Yao Lyu, Jie Cai, Anisa Callis, Kelley Cotter, and John M Carroll. " I Got Flagged for Supposed Bullying, Even Though It Was in Response to Someone Harassing Me About My Disability.": A Study of Blind TikTokers' Content Moderation Experiences. In *Proc. CHI*, 2024.

[64] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. Reliability and Inter-Rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.*, 2019.

[65] Ashlee Milton, Leah Ajmani, Michael Ann DeVito, and Stevie Chancellor. "I See Me Here": Mental Health Content, Community, and Algorithmic Curation on TikTok. In *Proc. CHI*, 2023.

[66] Azadeh Mokhberi, Yue Huang, Guillaume Humbert, Borke Obada-Obieh, Masoud Mehrabi Koushki, and Konstantin Beznosov. Trust, Privacy, and Safety Factors Associated with Decision Making in P2P Markets Based on Social Networks: A Case Study of Facebook Marketplace in USA and Canada. In *Proc. CHI*, 2024.

[67] Angela Molem, Stephann Makri, and Dana Mckay. Keepin' It Reel: Investigating How Short Videos on TikTok and Instagram Reels Influence View Change. In *Proc. CHIIR*, 2024.

[68] Seung Ho Na, Sumin Cho, and Seungwon Shin. Evolving Bots: The New Generation of Comment Bots and Their Underlying Scam Campaigns in Youtube. In *Proc. IMC*, 2023.

[69] Meera Navlakha. These TikTok scams are very real. https://mashable.com/article/tiktok-scams, October 2022. Accessed: 2025-06-02.

[70] Rajvardhan Oak and Zubair Shafiq. Victims, Vigilantes, and Advice Givers: An Analysis of Scam-Related Discourse on Reddit. In *Proc. SOUPS*, 2025.

[71] Rajvardhan Oak and Zubair Shafiq. "Hello, Is This Anna?": Unpacking the Lifecycle of Pig-Butchering Scams. In *Proc. SOUPS*, 2025.

[72] Tim Owen, Wayne Noble, and Faye Christabel Speed. The Challenges Posed by Scammers to Online Support Groups: The 'Deserving' and the 'Undeserving' Victims of Scams. *New perspectives on cybercrime*, 2017.

[73] Katalin Parti and Faika Tahir. "If We Don't Listen to Them, We Make Them Lose More than Money:" Exploring Reasons for Underreporting and the Needs of Older Scam Victims. *Social Sciences*, 2023.

[74] Gordon Pennycook, Adam Bear, Evan T Collins, and David G Rand. The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. *Management Science*, 2020.

[75] Katharina Pfeffer, Alexandra Mai, Edgar Weippl, Emilee Rader, and Katharina Krombholz. Replication: Stories as Informal Lessons About Security. In *Proc. SOUPS*, pages 1–18, 2022.

[76] Thomas Poell, David B Nieborg, and Brooke Erin Duffy. *Platforms and Cultural Production*. John Wiley & Sons, 2021.

[77] Emilee Rader and Janine Slaker. The Importance of Visibility for Folk Theories of Sensor Data. In *Proc. SOUPS*, 2017.

[78] Emilee Rader, Rick Wash, and Brandon Brooks. Stories as Informal Lessons About Security. In *Proc. SOUPS*, 2012.

[79] Reddit user. Sugar babies of reddit, what are your stories and experiences? https://doi.org/10.5281/zenodo.20707051, 2016.

[80] Reddit user. So how does this reddit sugar baby scam work? https://doi.org/10.5281/zenodo.20633128, 2021.

[81] Elissa M Redmiles, Sean Kross, and Michelle L Mazurek. How I Learned to Be Secure: A Census-Representative Survey of Security Advice Sources and Behavior. In *Proc. CCS*, 2016.

[82] Isma Azis Riu and Irsan Radjab. Influencer Marketing on Tiktok: A Literature Review Analysis. *Journal of Public Administration and Government*, 2023.

[83] Fatima Salahdine and Naima Kaabouch. Social Engineering Attacks: A Survey. *Future Internet*, 2019.

[84] Patrawat Samermit, Anna Turner, Patrick Gage Kelley, Tara Matthews, Vanessia Wu, Sunny Consolvo, and Kurt Thomas. "Millions of People Are Watching You": Understanding the Digital Safety Needs and Practices of Creators. In *Proc. USENIX*, 2023.

[85] Dawn M Sarno and Jeffrey Black. Who Gets Caught in the Web of Lies?: Understanding Susceptibility to Phishing Emails, Fake News Headlines, and Scam Text Messages. *Human Factors*, 2024.

[86] Joseph S Schafer, Annie Denton, Chloe Seelhoff, Jordyn Vo, Lance Garcia, Isha Madan, Alisha Mudbhary, Ruijingya Tang, and Kate Starbird. 'I Blow up': Understanding TikTok Users' Reactions to Sudden Social Media Attention. *Proc. ACM Hum.-Comput. Interact.*, 2025.

[87] Jason Scharfman. Crypto romance scams and pig butchering. In *The Cryptocurrency and Digital Asset Fraud Casebook, Volume II: DeFi, NFTs, DAOs, Meme Coins, and Other Digital Asset Hacks*. Springer, 2024.

[88] Filipo Sharevski, Jennifer Vander Loop, Peter Jachim, Amy Devine, and Emma Pieroni. User Experiences with Abortion Misinformation on TikTok: Encounters, Assessment Strategies, and Response. In *Proc. CHI*, 2025.

[89] Gilberto Atondo Siu, Alice Hutchings, Marie Vasek, and Tyler Moore. "Invest in Crypto!": An Analysis of Investment Scam Advertisements Found in Bitcointalk. In *2022 APWG Symposium on Electronic Crime Research (eCrime)*, 2022.

[90] Frank Stajano and Paul Wilson. Understanding Scam Victims: Seven Principles for Systems Security. *Commun. ACM*, 2011.

[91] Ella Steen, Kathryn Yurechko, and Daniel Klug. You Can (Not) Say What You Want: Using Algospeak to Contest and Evade Algorithmic Content Moderation on TikTok. *Social Media+ Society*, 2023.

[92] Jason Striegel. TikTok demographics 2024: Age & gender distribution. https://www.blogcadre.com/resources/tiktok-demographics/, 2024. Accessed: 2025-06-04.

[93] Karthik Subbian, Dhruv Sharma, Zhen Wen, and Jaideep Srivastava. Social Capital: The Power of Influencers in Networks. In *AAMAS*, 2013.

[94] Kurt Thomas, Patrick Gage Kelley, Sunny Consolvo, Patrawat Samermit, and Elie Bursztein. "It's Common and a Part of Being a Content Creator": Understanding How Creators Experience and Cope with Hate and Harassment Online. In *Proc. CHI*, 2022.

[95] TikTok. Oxford economics reports: Smb's use of tiktok contributed $24.2 billion to u.s. economy in 2023. https://newsroom.tiktok.com/tiktok-economic-impact-report-2024-smb?lang=en, 2024. Accessed: 2026-02-19.

[96] TikTok. Frauds and scams. https://www.tiktok.com/safety/en/scams, 2025. Accessed: 2025-04-15.

[97] TikTok. Tiktok for developers - research tools, 2025. https://developers.tiktok.com/products/research-api/.

[98] TikTok Shop. TikTok Shop Seller Center. https://business.tiktokshop.com/us/seller, 2025. Accessed: 2025-06-04.

[99] TikTok Shop Partner Center. Multi-Channel Network (MCN). https://partner.tiktokshop.com/docv2/page/650a39fe0fcef602bf26c0d3, 2025. Accessed: 2025-05-29.

[100] TikTok Support. How to Tell if an Account is Verified on TikTok. https://support.tiktok.com/en/using-tiktok/growing-your-audience/how-to-tell-if-an-account-is-verified-on-tiktok, 2025. Accessed: 2025-06-04.

[101] Raj Vardhan, Alok Chandrawal, Phakpoom Chinprutthiwong, Yangyong Zhang, and Guofei Gu. # DM-Me: Susceptibility to Direct Messaging-Based Scams. In *Proc. ASIACCS*, 2023.

[102] Julia Velkova and Anne Kaun. Algorithmic Resistance: Media Practices and the Politics of Repair. *Information, Communication & Society*, 2021.

[103] Julie A. Vera and Sourojit Ghosh. "They've Over-Emphasized that One Search": Controlling Unwanted Content on TikTok's for You Page. In *Proc. CHI*, 2025.

[104] Fangzhou Wang. Breaking the Silence: Examining Process of Cyber Sextortion and Victims' Coping Strategies. *International Review of Victimology*, 2025.

[105] Fangzhou Wang and Volkan Topalli. Understanding Romance Scammers Through the Lens of Their Victims: Qualitative Modeling of Risk and Protective Factors in the Online Context. *American Journal of Criminal Justice*, 2024.

[106] Fangzhou Wang and Xiaoli Zhou. Persuasive Schemes for Financial Exploitation in Online Romance Scam: An Anatomy on Sha Zhu Pan in China. *Victims & Offenders*, 2023.

[107] Rick Wash. Folk Models of Home Computer Security. In *Proc. SOUPS*, 2010.

[108] Rick Wash and Molly M. Cooper. Who Provides Phishing Training? Facts, Stories, and People Like Me. In *Proc. CHI*, 2018.

[109] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-Of-Thought Prompting Elicits Reasoning in Large Language Models. *Advances in Neural Information Processing Systems*, 2022.

[110] Miranda Wei, Sunny Consolvo, Patrick Gage Kelley, Tadayoshi Kohno, Tara Matthews, Sarah Meiklejohn, Franziska Roesner, Renee Shelby, Kurt Thomas, and Rebecca Umbach. Understanding Help-Seeking and Help-Giving on Social Media for Image-Based Sexual Abuse. In *Proc. USENIX*, 2024.

[111] Miranda Wei, Eric Zeng, Tadayoshi Kohno, and Franziska Roesner. Anti-Privacy and Anti-Security Advice on TikTok: Case Studies of Technology-Enabled Surveillance and Control in Intimate Partner and Parent-Child Relationships. In *Proc. SOUPS*, 2022.

[112] Mika Westerlund. The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, 2019.

[113] Monica T. Whitty. Do You Love Me? Psychological Characteristics of Romance Scam Victims. *Cyberpsychology, Behavior, and Social Networking*, 2018.

[114] Monica T Whitty and Tom Buchanan. The Online Romance Scam: A Serious Cybercrime. *CyberPsychology, Behavior, and Social Networking*, 2012.

[115] Justin Wu and Daniel Zappala. When Is a Tree Really a Truck? Exploring Mental Models of Encryption. In *Proc.*, 2018.

[116] Qunfang Wu, Yisi Sang, Dakuo Wang, and Zhicong Lu. Malicious Selling Strategies in Livestream E-Commerce: A Case Study of Alibaba's Taobao and ByteDance's TikTok. *ACM Trans. Comput.-Hum. Interact.*, 2023.

[117] Pengcheng Xia, Haoyu Wang, Bowen Zhang, Ru Ji, Bingyu Gao, Lei Wu, Xiapu Luo, and Guoai Xu. Characterizing Cryptocurrency Exchange Scams. *Computers & Security*, 2020.

[118] Yuxiang Zhai, Xiao Xue, Zekai Guo, Tongtong Jin, Yuting Diao, and Jihong Jeung. Hear Us, Then Protect Us: Navigating Deepfake Scams and Safeguard Interventions with Older Adults Through Participatory Design. In *Proc. CHI*, 2025.

[119] Zeyuan Zuo, Ningxin Su, Baochun Li, and Teng Zhang. PACK: Towards Communication-Efficient Homomorphic Encryption in Federated Learning. In *Proc. SoCC*, 2024.

A Appendix

A.1 Scam-type Example Gallery

Examples of various scam-types[10]:

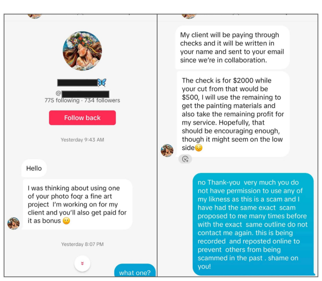


Figure 3: Example mural scam.

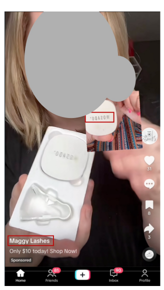
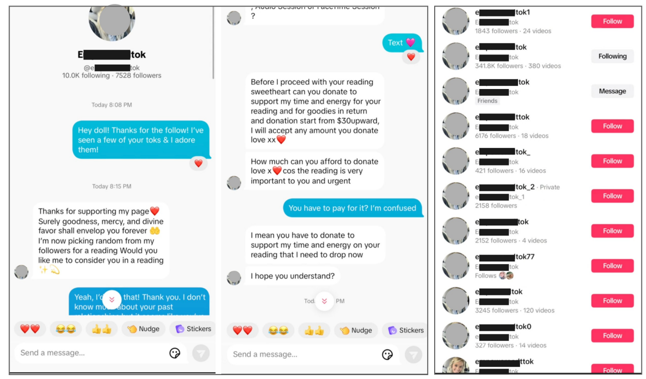


Figure 4: AI-driven deepfake of a creator

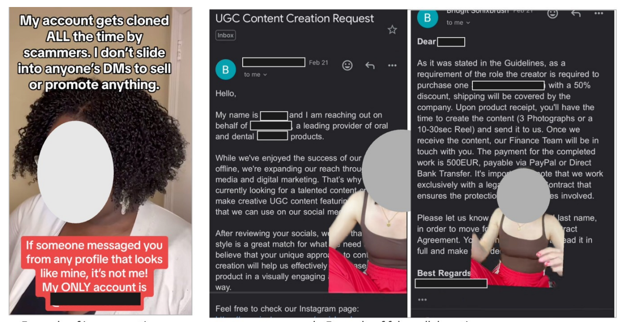
[10] Scam-type Examples: https://osf.io/yqxb9/files/m7ftb



a. Example of psychic reading scam

b. Example of profile impersonation

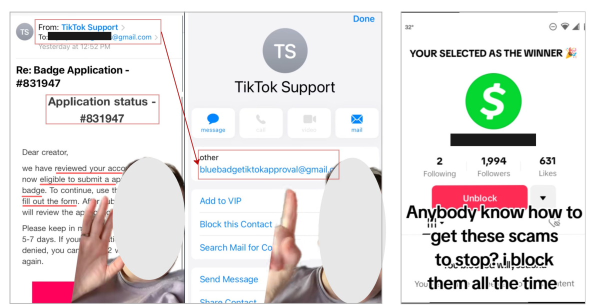
Figure 5: Overlap between psychic reading and impersonation scams: T18 found out multiple accounts of the psychic reader.



a. Example of impersonation scam

b. Example of fake collaboration scam

Figure 6: Examples of a) Impersonation scam, and b) Fake collaboration scam.



a. Example of account verification badge scam

b. Example of tagging user scam

Figure 7: Examples of account-related scams.

Table 2: Defining scam types and mapping them to prior work

Types of scams	Definition and example in the context of TikTok	Related Scams in Prior Work and Difference
Finance scams (54/33%): Lure people into financially rewarding schemes that turn out to be bogus.		
Mural	Scammer poses as an artist and requests eCheck payments for murals. E.g., T13 received a DM offering to paint her for a payment of $2500 via eCheck and was asked to send $2000 back to the "artist" for supplies. She declined. This was also reported by T70 via Instagram DM	Fake check / fake buyer scams [9]. Reconfigures fake-check overpayment model into SVP-specific variation involving artist-impersonation scheme.
Psychic reading	Claim to tell a victim's fortune for a fee.	Fortune-telling fraud [9].
Crowdfunding	Crowdfunding for bogus causes (e.g., fake illness). E.g., T67 encountered a post requesting funds to treat a sick cat. She felt bad and made a donation via CashApp.	Charitable solicitation scams [9]. Expands charitable solicitation to include personal appeals (e.g., sick pets, pending surgery) requested via GoFundMe links in DMs/FYP on TikTok
Fake investment opportunity	Lure victims into easy-money schemes.	Misc investments, Ponzi schemes, Crypto scams [9, 70].
Scams targeting influencers (49/30%): Targets content creators through account impersonations and fake opportunities.		
Impersonation	Scammers create fake profiles of creators and reach out to their followers to request money. E.g., T31: "*My account gets cloned all the time by scammers. I don't slide into anyone's DM to sell anything (paraphrased).*"	Imposter Scams [9, 37, 70]. Shifts impersonation from institutional authority (tech support) or family and friends (distressed relative) to creators, exploiting influencer-centric culture and a para-social relationship with followers.
Fake collaboration	Companies request creators to endorse products, but require them to purchase those products first. E.g, T145: *"I was hyped for my first offer until I realized it was the same scam (paraphrased)."*	Business opportunity fraud [9], Employment/job fraud [36, 70]. Shifts focus from a general audience to creators, leveraging their desire to grow on TikTok using a pay-to-avail model.
Social media brand manager	Targets creators in early growth stages with promises to increase followers and view counts. E.g, A scammer contacted T61, offering opportunities to grow account, but the agency's TikTok account was banned.	Business coaching scams [9], Sports agent and endorsement scams [36], Fake LinkedIn recruiters [70]. Differs by reconfiguring fee-based professional growth scams into talent agencies for growing TikTok creators.
Online shopping scams (36/22%): Fraudulent experiences when purchasing products or services online.		
Fake or poor-quality product	False or misleading information about products that are worth much less.	Worthless products [9], Fake websites, e-commerce scams, Counterfeit/shoddy product scams [54, 66, 70].
Undelivered goods	Payment made for a product that is never received.	Paid never received [9].
Unauthorized billing	Charged for products or services not agreed to.	Unauthorized billing [9], Fake subscription/billing [70].
Account-related scams (17/10%): Users are tricked into giving away personal information.		
Tag users in contests	Users are tagged in false prize offerings. E.g., T60 shared being publicly tagged in posts that used the CashApp logo as the profile picture to appear legitimate.	Prize, sweepstakes, and lottery scams [36]. Differs by use of platform affordances like public tagging and payment-app logos (e.g., CashApp) to appear legitimate.
Account badge verification	Fake badge verification offers to creators from scammers impersonating TikTok. E.g., T77 found an official-looking email with a TikTok logo and a badge application number.	Not found. Combines institutional impersonation [9] of TikTok and phishing [36] to exploit the creator economy.
Account hack	User accounts compromised through deception.	Family & Friend Imposters [9]. Used with above scam.
Romance scams (10/6%): Emotional manipulation for financial gain.		
Traditional romance scam	Fake romantic interest used to extort money.	Romance scams [9, 113, 114].
Sugar daddy/baby scam	A wealthy person offers money initially, then requests favors or payments. E.g., T64 reported being contacted via DM, *"He offered to send money, then asked me to move to Google Chat. (paraphrased)"*.	Romance scams [9, 113, 114], Pig-butchering scams [70, 71], Reddit [80]. A variant of romance scam adapted to SVPs, often targets young women, promising financial companionship at the outset rather than gradually cultivating romantic attachment.
Miscellaneous scams (5/3%)		
AI-driven	Use of AI-generated deepfakes. E.g., Scammers used an AI-altered video of T98 to falsely promote a product.	Deepfake-enabled fraud [112, 118]. Focuses on synthetic creator endorsements rather than deepfakes targeting older adults.
Health-related	Fake health products and fitness programs.	Weight-loss products, health supplements [9].
Educational financial	Scams offering fake educational financial support.	Scholarship or educational grant scams [9].

Table 3: LLM Prompt to perform two-level classification of the comment texts. Note that our analysis leverages the Supportive comment type.

Role
You are an expert Social Media Researcher specializing in digital safety, scams, and community discourse. Your task is to classify TikTok comments based on their relationship to the video's content and the creator's intent.

Task
Analyze the provided TikTok video description and the comment. You will perform a two-step classification based on the criteria below.

Classification Criteria

Level 1: Stance (Alignment with Creator's Message)

1. **Supportive**: The comment aligns with the creator's stance or the video's theme of scam awareness/prevention. This includes:
   - **Validation/Explanation**: Agreeing with a warning or explaining the *mechanism* of the scam (e.g., mentioning "dropshipping" or "affiliate links").
   - **Solicitation/Recovery**: Offering "help," "recovery services," or tagging "recovery experts" (e.g., "Contact @User for a refund"). Even if these appear to be automated bots, they are categorized as Supportive because they align with the "scam resolution" theme.
   - **Engagement**: Expressing gratitude, sympathy, or asking follow-up questions about the scam.
2. **Against**: The comment challenges, contradicts, or mocks the creator's message. This includes:
   - Defending a reported scammer or product (e.g., "It arrived for me").
   - Dismissing the creator's concern as "fake" or "exaggerated."
3. **Neutral**: The comment is a surface-level observation or functionally independent. This includes:
   - **Reactionary/Humorous**: Reacting to a specific funny visual or audio moment without commenting on the scam itself (e.g., "I lost it at the [visual element]").
   - **Observations/Lists**: Listing names or quoting prices mentioned in the video without providing an opinion (e.g., "I saw [person] too").
   - **Unrelated Content**: Spam emojis or questions about unrelated topics (e.g., "Where is your shirt from?").

Level 2: Supportive Sub-type (Apply ONLY if Level 1 is 'Supportive')

- **Useful**: The user indicates a knowledge transfer. They mention they "didn't know this," "learned something," or provide additional technical context that helps others understand the scam.
- **Relatable**: The user shares a personal anecdote or confirms they have experienced the same thing (e.g., "This happened to me," "I saw this too").
- **Neither**: General agreement, praise, or bot-like solicitations that lack a specific personal story or a statement of learning.
- **N/A**: Use only if Level 1 is 'Against' or 'Neutral'.

Rules for Classification

1. **Mechanism = Support**: If a comment identifies the business model of the scam (e.g., "This is just dropshipping"), classify as **Supportive**.
2. **Recovery Bots = Support**: Comments claiming a third party can "recover money" or "is a life saver" regarding scams are **Supportive** (Alignment with the scam theme).
3. **Surface Reactions = Neutral**: Quoting a funny moment or stating a price mentioned in the video without a stance is **Neutral**.
4. **Context is King**: A "mean" tone doesn't always mean 'Against.' If the user is angry at the scammer, that is **Supportive** of the creator.

Output Format (Strict JSON)

```
{
 "analysis": "1. Creator's point? 2. Comment stance (Does it explain, react, or contradict)? 3. Personal story or learning?",
 "level_1_stance": "Supportive | Against | Neutral",
 "level_2_subtype": "Useful | Relatable | Neither | N/A"
}
```

Examples
...

Input Data
Description: {video_description}
Comment: {comment}

Classification Result

Table 4: The values of engagement metrics in referenced videos.

Participant	Video Type	Type of Scam	Views	Likes	Comments	Shares	Example Comments (Paraphrased)
T10	Reporting and advising	Online shopping	49	35	6	0	*"Mine never came! I hadn't realized the sale happened outside of TikTok Shop."*
T18	Scam reporting	Finance	17	21	11	0	*"I hate how scammy TikTok has become. I rarely trust or talk to strangers here."*
T22	Anti-scam advice	Finance	35	7	3	0	*"I had the same issue until they helped me fix it quickly."*
T23	Scam reporting	Influencer targeting	22	40	12	0	*"This has gone on forever. it's insane that creators can't stop it."*
T30	Scam reporting	Online shopping	2	2	0	0	
T31	Scam reporting	Influencer targeting	48	32	2	0	*"Clones usually copy my entire profile and content but use a slightly misspelled username."*
T39	Reporting and advising	Finance	14	38	16	0	
T43	Scam reporting	Finance	67	93	15	0	
T45	Scam reporting	Finance	4	22	19	0	*"I can't believe these are scams?!"*
T61	Reporting and advising	Influencer targeting	56	23	7	0	
T62	Scam reporting	Online shopping	10	10	6	0	*"They charged my card the day after starting the trial, so I disputed it."*
T64	Scam reporting	Finance and romance	115	43	19	0	*"I'm so over it. No, I don't want a mural and I don't want your 'free' money."*
T70	Scam reporting	Finance	3	19	5	0	*"Total scam. I just block anyone with typos and a bunch of photos posted all at once."*
T73	Scam reporting	Influencer targeting	2	8	3	0	*"They blocked me, but I'm still hearing from people they are trying to scam."*
T77	Scam reporting	Account	3121	630	120	0	*"The scammers are working overtime! Same thing happened to me on X, glad you're back."*
T95	Anti-scam advice	Account	308	307	15	0	*"I accidentally gave my login details to a scammer. can I still save my account?"*
T97	Reporting and advising	Influencer targeting	29	20	5	0	*"I already sent them a nasty reply calling them out. hope karma hits them hard."*
T98	Scam reporting	Online shopping	261	57	13	0	*"They got me. I thought it was you, but it was a scam."*
T100	Scam reporting	Influencer targeting	12432	8278	1070	9	*"I saw a video about a girl who never got her commissions and only got excuses."*
T126	Scam reporting	Finance	0	0	0	0	
T145	Scam reporting	Influencer targeting	530	156	85	13	*"I was hyped for my first offer today until I realized it was the same scam."*