“Is this a scam?”: The Nature and Quality of Reddit Discussion about Scams

Elijah Bouma-Sims
eboumasi@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Mandy Lanyon
mandy@cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Lorrie Faith Cranor
lorrie@cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Abstract

People often use social media platforms to seek advice about scams like ecommerce fraud or phishing; however, little research has investigated the nature of such discussion. We conducted a multi-stage thematic analysis of 1,525 posts made to four communities focused on scam discussion on Reddit, primarily from `/r/Scams`. We found that posters use Reddit to identify scams, discuss the strategies employed by scammers, and obtain advice on coping with victimization. The scams discussed are primarily mediated by the internet or related technologies. Users in the communities we studied especially provide informational support and reassurance to victims, although some comments reinforce victim-blaming attitudes. We also observed qualitative differences in the types of support sought and given based on the community, with the board `/r/Sextortion` especially being used for emotional support. We conclude that Reddit's scam discussion communities serve as a valuable resource for scam prevention and remediation. Additionally, we discuss the potential for future research and law enforcement engagement on Reddit.

CCS Concepts

• Social and professional topics → Computer crime; **• Security and privacy → Social aspects of security and privacy**; **• Human-centered computing → Empirical studies in collaborative and social computing**.

Keywords

Scams, Reddit, Social media, Security advice

ACM Reference Format:
Elijah Bouma-Sims, Mandy Lanyon, and Lorrie Faith Cranor. 2025. “Is this a scam?”: The Nature and Quality of Reddit Discussion about Scams. In *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security (CCS '25), October 13–17, 2025, Taipei, Taiwan.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3719027.3765030

1 Introduction

Phishing, fraudulent ecommerce websites, financial sextortion, and other deceptive schemes (collectively referred to as “scams”) are the most common internet safety threats faced by end-users [9, 23, 25]. Many scams go unreported [3, 33, 56], but estimates from the Global Anti-Scam Alliance suggest that users lost $1 trillion to scams in 2024 [1], with much of this money being unrecoverable. Left to cope with a dizzying array of threats, many users seek out online sources to help avoid scams or recover from victimization [11, 12, 31, 44, 69, 70]. The American social media website Reddit, in particular, has a number of communities focused on the discussion of scams, with the largest, `/r/Scams`, having nearly a million members as of February 2025.

Despite the prevalence of scam-related online help-seeking, there is a lack of research characterizing general social media discussion about scams [12, 44], with much of the existing work focusing on financial sextortion [11, 69, 70]. To address this gap, we conducted a multi-stage thematic analysis of a sample of 1,525 posts and 1,883 associated comments from four communities centered on scam discussion on Reddit, with the majority of posts coming from `/r/Scams`. We sought to answer the following research questions:

(1) What types of scams are discussed on Reddit?
(2) What types of scam-related support do users seek on Reddit?
(3) What security advice or other help do victims of online scams receive on Reddit and what is the quality of this advice?

We found that Reddit discussion focuses on internet-mediated scams that constitute crimes. Discussion on Reddit focused on identifying scams, discussing the strategies employed by scammers, and requesting advice to deal with victimization. In addition to seeking general remediation advice, users also commonly sought reassurance that they were safe from harm and other emotional support. We noted variance between communities, with posters especially using `/r/Sextortion` to give and receive emotional support.

We conclude that Reddit serves as a valuable resource that helps users avoid scams and cope with victimization. Still, the prevalence of scam-related discussions highlights the difficulty that users have in avoiding scams online. We encourage the development of automated tools to help users reason about scams, discuss the possibility of future research based on Reddit data, and suggest increased law enforcement outreach to Reddit's scam-related communities.

2 Background and Related Work

“Scam” is an informal term used both colloquially and in the academic literature. While usages vary, we adopt the definition from Button and Cross: a scam is a “deceptive scheme that seeks to trick a person(s) out of money and/or personal information...” [10, pg.7]. In particular, we focus on internet scams, which are initiated through the internet or related communications technologies (e.g., SMS, phone calls, social media, etc.) and use the term “scam” to refer to internet scams for the remainder of this paper. Scams are among the most common cybercrimes. For example, phishing, in which an attacker impersonates a trusted entity to obtain sensitive private information such as a user's password or tax ID number,

CCS '25, Taipei, Taiwan

ACM ISBN 979-8-4007-1525-9/2025/10
https://doi.org/10.1145/3719027.3765030

was the single most common crime reported to the Federal Bureau of Investigation (FBI) Internet Crime Complaint Center (IC3) in 2023 [23]. Due to its significance as a threat to organizational and individual security, phishing has received the most attention from researchers [18, 47, 74].

Many of those victimized by scams do not seek help from governmental or personal sources. Prior studies have generally found that most scam cases go unreported [3, 33, 56]. For example, an analysis of surveys administered by the FTC in 2005, 2011, and 2017 suggests that only 4.8% of U.S. mass-marketing scam victims reported their experience to a government agency or the Better Business Bureau (BBB) [3]. Alongside other factors [37], under-reporting may result from the pervasive stigma around victimization. Scam victims are often viewed as responsible for their experience and maligned as gullible [13, 16].

In this context, social media can be a valuable tool in helping scam victims recover, allowing them to receive personalized support without revealing their real identity. Our study is based on posts and comments on the American social media platform, Reddit. The platform is composed of thousands of user-created and moderated communities called "subreddits," that focus primarily on specific topics (e.g., `/r/NFL` for the National Football League in the United States) or identity groups (e.g., `/r/TwoXChromosomes` for women). Users submit posts that link to external websites or contain original content. Users vote to determine the default order of posts, with an "upvote" generally boosting a post and a "downvote" lowering the post's visibility. Users can also comment on posts, with the order of comments determined by the same voting scheme. Reddit has a number of communities dedicated to receiving advice, including advice related to scams (e.g., `/r/Scams` and `/r/Sextortion`). Reddit is a particularly popular source for research due to its community structure and the relative accessibility of data [55]. Reddit's user base is skewed towards young adults, with 46% of Americans aged 18 to 29 using the service, as compared to only 4% of adults 65 or older [29].

Prior work in security and privacy has analyzed online discussion from Reddit and other social media platforms to explore a variety of topics, such as software development practices [42, 65, 66, 72] or end-user attitudes towards security and privacy [2, 41, 53]. Some studies have explored the general types of security-related support sought by non-experts online [7, 31]. For example, Hasegawa et al. analyzed 445 security-related questions from a Japanese Q&A, identifying the most common topics for questions [31]. Notably, they found that the largest category of questions concerned cyberattacks, particularly scams such as phishing, underscoring the prominence of scams in public-facing security discourse. Bhagavatula et al. analyzed 200,000 Twitter or Facebook posts from 38 participants, finding that constructive security and privacy advice was rare, with many relevant posts sharing dangerous security habits (e.g., password reuse) [7]. Other studies have examined more specific areas of security advice, such as discussion of the legality of port scanning on Reddit [34].

Only a small number of studies have investigated help-seeking related to scams on social media [11, 12, 44, 49, 69, 70]. Most relevant to the present study is a paper by Oak and Shafiq that analyzed 2,435 scam-related Reddit posts to examine how users conceptualize scams and the role of Reddit communities in prevention and support [49]. While our work shares some similar goals, the studies were conducted concurrently and independently. Oak and Shafiq apply different methodological choices (e.g., excluding image-only posts) and provide broader thematic analysis. Our study offers more fine-grained insight into the specific types of support users seek and receive in scam-related subreddits. A number of papers have examined Reddit discussion of financially-motivated sextortion [11, 69, 70], a type of scam where victims are tricked into sharing intimate imagery and then extorted with the threat of publication. Sextortion may also refer to cases where the scammer falsely claims to have explicit media featuring the victim [51]. For example, Wei et al. [70] used large-language models (LLMs) to identify over 100,000 posts across Reddit where users sought advice about image-based sexual abuse (IBSA), including financially-motivated sextortion. They analyzed a stratified sample of 261 posts (50 of which focused on financial sextortion) to evaluate how users seek and receive help with IBSA on Reddit. Focusing on scams involving crypto assets, Childs [12] qualitatively analyzed the top 200 posts from `/r/CryptoCurrency` that mentioned the word "scam" to assess how users prevent and cope with victimization. Our paper is the first to provide qualitative insight into general social media discussions and advice-seeking about scams.

Another line of research has focused on online vigilantism, referred to as "scambaiting," wherein users pose as victims and interact with scammers to waste their time or cause harm (e.g., by deleting files). While the practice of scambaiting has existed for decades [67], in recent years, it has become more prominent with popular scambaiting creators emerging on user-generated content websites like YouTube and Twitch [6, 59]. On Reddit, `/r/scambaiting` is dedicated to the practice, with users sharing examples of scams and their experiences interacting with scammers [20]. Most research on the practice has focused on characterizing the social dynamics of scambaiting [6, 59], without a focus on security and privacy issues. One exception is a study conducted by Edwards et al. [21], which utilized transcripts of email exchanges from scambaiting communities to measure the use of persuasion tactics used by scammers in advance-fee scams and train a classifier to distinguish between scammer and scambaiter emails. Due to the artificial nature of scambaiting interactions and the potential for wholly fabricated exchanges raised by Dynel and Ross [20], we exclude scambaiting content from our analysis.

3 Methods

We conducted a multi-stage thematic analysis of 1,525 posts from a set of Reddit communities dedicated to discussing scams.

3.1 Dataset

We did not collect data directly using Reddit's API. Rather, we used an archive of data collected via PushShift [4] from June 2005 through December 2022 and ArcticShift[1] from January to December 2023 (inclusive). This dataset includes the content of some posts deleted by users or removed by moderators, so it is more complete than the dataset available directly from Reddit. ArtcicShift is a successor to the widely used PushShift project, which is no longer generally

[1] https://github.com/ArthurHeitmann/arctic_shift

available due to changes in Reddit policy. Other recent research on Reddit has also used data from ArcticShift (e.g., see Shen et al. [63]).

For our content analysis, we selected four communities dedicated to discussing scams: /r/Scams, /r/Sextortion, /r/scammers, and /r/phishing. /r/Scams has nearly a million members as of February 2025 and is by far the largest community focused on scams on Reddit. /r/scammers is a much smaller community that also focuses on the general discussion of scams. /r/Sextortion and /r/phishing focus on specific types of scams, as indicated by their names. These communities were selected based on the authors' prior knowledge of Reddit and by entering the terms "scam," "fraud," and "phishing" in Reddit's community search.[2] We chose not to include large communities that seemed to focus on online vigilantism, referred to as "scambaiting," in which users pose as victims and interact with scammers to waste their time or cause other harm [20, 67] (e.g., /r/ScamBait and /r/ScamNumbers) due to the artificial nature of scambaiting interactions and the potential for wholly fabricated exchanges raised by Dynel and Ross [20].

We collected 278,231 posts and 2,302,800 associated comments from these communities made between June 14, 2008 (the date /r/phishing was created), and December 31, 2023. The content of 84,091 of these posts (30.2%) were deleted by the original poster or removed by moderators prior to being archived by PushShift or ArcticShift. Of the remaining 194,140 posts, we randomly sampled 1,525 (0.8%) with 15,240 associated comments. This sample size was chosen to achieve a 95% confidence level with a 2.5% margin of error. The final sample was composed of $1,311$ posts from /r/Scams (86.0%), 144 posts from /r/Sextortion (9.4%), 42 posts from /r/scammers (2.8%), and 28 posts from /r/phishing (1.8%).

3.2 Thematic Analysis

To investigate our research questions, two authors performed thematic analysis in multiple stages. They developed codebooks that can be found in the external Appendix.[3]

3.2.1 Stage 1: Initial Categorization. The first stage of coding was intended to characterize the type of content posted to the four selected scam-discussion communities to facilitate further analysis. To develop an initial codebook, the lead coder reviewed a subset of 100 posts from the sample and performed thematic analysis. The coder viewed the post title and post body. For non-spam posts, if a post contained a link to content that was not clearly irrelevant, the coder also considered this content. At this stage, the coder did not systematically analyze the comments, although we often read comments in part to aid in categorizing the post (e.g., when the poster provided additional information in the comments).

We primarily categorized posts based on their apparent purpose (e.g., *scam identification*, *request for support*, *report of victimization*, etc.). Many posts linked to images or videos to show the scam that was being discussed. If these were no longer available, we coded the post as "media missing" and did not review it during subsequent analysis (255 posts, 16.7% of the sample). For posts that discussed a scam, we characterized the scam based on the initial method used to communicate with the potential victim (e.g., social media, instant messaging, etc.) and other technologies used in the scam (e.g., crypto assets, gift cards, etc.). We also assigned a general "type" code that was meant to capture the distinctive characteristics of the scams (e.g., *relationship scam* for scams that involve a scammer pretending to be a friend or romantic partner, *ecommerce seller* for scams perpetrated by a purported online seller of goods/services, etc.). The specific codes we chose were informed by the language used by Reddit users and the terminology used in prior research. Finally, for posts where users reported victimization by scams, we assigned a code for any harm experienced due to the scam (e.g., financial loss, exposure of personal information, etc.).

After developing the initial codebook, the lead coder and another author collaboratively coded the full sample of posts. We dual-coded 400 posts, with the coders independently coding blocks of 100, comparing the results, and meeting to resolve any disagreements. During this phase, we added new codes as appropriate. On the fourth set of 100, the coders achieved Krippendorff's α [39, 45] of 0.951 for post type, 0.895 for initial communication method, 0.905 for technologies used, and 0.966 for type of harm experienced by a victim. Considering the time-consuming nature of dual coding the entire sample, we divided the remaining posts equally between the coders and coded separately. The coders met regularly throughout the separate coding process to discuss difficulties and emerging themes.

3.2.2 Stage 2: Analysis of Advice. The second coding stage was intended to analyze the experience of scam victims requesting support on Reddit and the type of advice they received. As such, we focused on the 262 posts (17.2%) that were coded as *requests for support* in the first stage.[4] For these posts, we identified the type of support that the victim requested, what remediation steps they had already taken (if any), what private information they had disclosed to the scammer (if any), and the magnitude of financial harm they had experienced (if any). At this stage, the comments on posts were also reviewed systematically to identify the support given to users. We read up to 15 comments per post, reading each thread from earliest to the latest. We chose this ordering, as we expect that victims are more likely to see and rely upon comments posted sooner rather than later. The default comment ordering shows the most-upvoted comments first, so this ordering does not reflect how other users viewing the post would experience the comments section. Where appropriate, we revisited the initial categorization of post type and scam type at this stage.

Coding followed a procedure similar to the first stage. The lead coder reviewed 50 posts and associated comments to develop an initial codebook. The two coders then dual-coded two sets of 50 posts and associated comments, meeting after each round of coding to compare results and resolve disagreements. On the second round of dual coding, the coders achieved Krippendorff's α of 0.822 for type of support that the victim requested, 0.829 for remediation steps the victim had already taken, 0.823 for type of private information disclosed, and 0.933 for comment codes. Considering the substantial agreement, the remaining posts and associated comments were divided evenly and coded separately.

[2] https://www.reddit.com/search/?type=communities

[3] https://doi.org/10.5281/zenodo.17048061

[4] Coding was initially performed with both posts coded as *report of victimization* and *request for support*. After codebook development, we decided to code only support requests.

3.2.3 Stage 3: Codebook refinement. After all posts and comments were assigned initial codes, the coders met to review and refine the initial codebook. In particular, the coders sought to group codes into larger categories and generate new codes that better captured the diversity of the data. Most notably, we grouped scam types into four nonexclusive categories, partially inspired by the categorization used in Deliema et al.'s investigation of the factors contributing to scam victimization [17]: impostor schemes involving impersonating a trusted entity (e.g., a business, the government, family/friends, etc.), fraudulent goods and services schemes premised on the sale of goods/services (e.g., non-delivery, non-payment, fake or defective products, etc.); opportunity schemes where the victim is attracted by the promise of something desirable (e.g., a prize, employment opportunity, etc.), and threat-based schemes where the victim is induced to pay money based on a threat of harm (e.g., arrest, release of intimate imagery, etc.). We also reviewed comments initially categorized as "Other" to determine what commonalities existed, and what new codes should be created. The final codebook includes 9 post-type codes and 39 scam-type codes.

After discussing changes to the codebook, the lead coder reviewed content marked with any of the codes that were split or merged (e.g., comments/posts that were marked with one of the "Other" codes). Any errors that the lead coder noticed were also corrected.

3.3 Ethical Considerations

Our research is not required to undergo IRB review, as we relied on publicly available data. Nevertheless, it is important to consider the ethical issues associated with social media research. Most users do not expect their posts to be analyzed as part of academic research [27], and we did not obtain consent from community members or moderators prior to conducting our research. In line with the recommendations of Fiesler et al., we did consider the privacy norms of the community before conducting our research [28]. As detailed in the results, many users on /r/Scams explicitly acknowledged the public nature of their posts and expressed a desire to help others by sharing their experiences. Members on /r/Sextortion discussed and shared research that was based on posts to their community [11]. We believe this context supports the ethical use of these posts in research. We commit to sharing our findings with the studied communities in an accessible manner once the paper is publicly available so that the studied communities may benefit from our findings.

We chose to include deleted posts and comments. We felt that excluding removed content may introduce bias if posts about certain types of scams are disproportionately deleted. Since this content is available via public archives (i.e., ArcticShift) and we are not resharing it, we feel that our analysis does not do additional harm. We only discuss deleted content in the aggregate and do not individually reference posts or comments that were deleted as of July 9th, 2025. Some comments on deleted posts are discussed, where the highlighted content remains available on Reddit.

Victims of crimes are a vulnerable population that may be targeted by criminals for revictimization (see Section 4.4 for a discussion of the scams targeting victims on Reddit). Moreover, scam victimization is highly stigmatized [13, 16] and users may not want to be publicly known as survivors of a scam. We take several steps to prevent the identification of the users we observed. First, we avoid mentioning the usernames of posters or commenters. Second, we will not publicly post our curated dataset, although the authors will provide an archived copy of the data upon reasonable request. Finally, most examples of posts or comments are described rather than quoted directly. Summarizing quotes was performed by editing original excerpts to replace original text with synonyms or alternate phrases that express the same idea. Some nouns were replaced with more general descriptors (e.g., a word like "cousin" or "brother" may become "family-member"). As recommended by Reagle [57], the lead author verified that Google or Reddit searches including the provided text did not return the highlighted post or comment.

There are two exceptions where we quote verbatim. We directly quote comments from the AutoModerator as these messages are not authored by individual users, and the excerpts cannot be associated with specific posts. The title includes a quote that was used by several users in scam identification posts. As this quote cannot be tied to any single user, we leave it verbatim.

3.4 Limitations

We cannot verify the claims made by users, and some posts and comments may contain intentional or unintentional falsehoods (e.g. when a user misunderstands what has happened to them). Further, we reviewed only a subset of communities centered on the discussion of scams, and users also seek advice about scam victimization on other communities (e.g., /r/Advice, /r/personalfinance, etc.). The nature of discussions and quality of advice may differ in these communities. Our sample skews towards /r/Scams. This subreddit accounts for the majority of Reddit posts about scams, but sampling randomly limits comparative analysis across subreddits. We also exclusively focused on comments and posts in English, so non-English speaking users are not represented in our sample. Finally, qualitative coding is inherently subjective, and a different team may have identified different themes in the same data.

4 Results

In this section, we describe the results of our content analysis and evaluation of advice quality. We begin with an overview of the sample before discussing results with respect to each research question.

4.1 Overview

Figure 1 shows the frequency of posts in the sample over time. In line with the growth of Reddit, the majority of posts (55.0%) in the sample come from 2022 and 2023. The median number of comments per post was 5, with 170 posts (11%) receiving no comments. Table 1 shows the breakdown of the different categories of posts, broken down by subreddit. We found that Reddit is used primarily to help users identify scams, discuss scammers' tactics, and seek support to deal with scams.

4.1.1 Scam identification. Across all subreddits, a plurality of the posts we reviewed (439 posts or 23.8%) were seeking to identify whether something was a scam (*scam identification*). These posts featured an image of communications from a potential scammer

Table 1: An overview of the types of posts we observed. Percentages refer to the proportion of posts in each column assigned to each type code. Proportions for subreddits other than `/r/Scams` should be interpreted with caution due to the small sample size.

Post Type	`/r/Scams`	(%)	`/r/Sextortion`	(%)	`/r/scammers`	(%)	`/r/phishing`	(%)	Sample	(%)
Scam identification	430	(32.8%)	2	(1.4%)	2	(4.8%)	5	(17.9%)	439	(28.8%)
Disc. of scam comms.	306	(23.3%)	2	(1.4%)	14	(33.3%)	4	(14.3%)	326	(21.4%)
Request for support	163	(12.4%)	93	(64.6%)	3	(7.1%)	3	(10.7%)	262	(17.2%)
Report of victimization	76	(5.8%)	26	(18.1%)	5	(11.9%)	1	(3.6%)	108	(7.1%)
Scambaiting	34	(2.6%)	3	(2.1%)	6	(14.3%)	1	(3.6%)	44	(2.9%)
Scam-related media	35	(2.7%)	1	(0.7%)	2	(4.8%)	4	(14.3%)	42	(2.8%)
Irrelevant	28	(2.1%)	4	(2.8%)	1	(2.4%)	2	(7.1%)	35	(2.3%)
Other	8	(0.6%)	3	(2.1%)	0	(0 %)	3	(10.7%)	14	(0.9%)
Media missing	231	(17.6%)	10	(6.9%)	9	(21.4%)	5	(17.9%)	255	(16.7%)

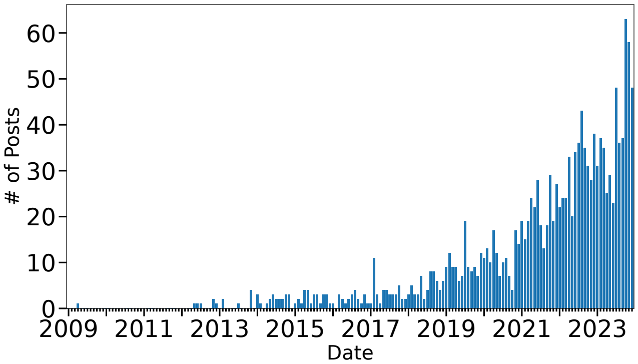

Figure 1: Frequency of posts in the sample over time.

or a description of an interaction that the poster was concerned about. In a typical example, a user made a post to `/r/Scams` asking whether a message they received was a scam. The post linked to an image of an SMS that threatened the recipient with arrest. The user received reassuring replies, with the first commenter stating that the government never sends messages threatening arrest and that the poster could safely ignore it. This result is similar to Hasegawa et al.'s study of non-expert security and privacy questions on a Japanese Q&A website (Yahoo! Chiebukuro), which found that the most common type of question sought to determine whether a user was facing a cyberattack [31].

In most cases, users posted scam identification questions before victimization. Only 23 posts in this category (5.2%) described some financial harm, and 19 posts in this category (4.3%) described the disclosure of personal information or intimate imagery to a scammer.

4.1.2 Discussion of scam communications. The next most common category was *discussion of scam communications*, with 326 posts (21.4%) assigned this code. This category is diverse, including any post in which users describe a scam they encountered or heard about but were not victimized by. Some of these posts were intended to warn others about a scam the user encountered. For example, one user made a post to `/r/Scams` where they described a phone call they received that attempted to collect on a fake debt. In explaining why they were posting, the user stated that they could not find a reference to the scam on the subreddit, so they posted to warn others. Posts designed to warn others about security threats were also noted in Bhagavatula et al.'s study of discussion of security and privacy on Twitter and Facebook, including posts warning about compromised accounts and a post about a Trojan being advertised on Facebook [7].

In some cases, users wanted to learn more about how a scam operated. In a typical example, one user made a post to `/r/Scams` where they asked about the purpose of a seemingly fraudulent job ad offering hundreds of dollars per day for less than eight hours of work.

Other posts in this category were intended to make fun of scammers' poor attempts at deception. For example, a user made a post to `/r/Scams` that linked to a fake receipt that was likely an attempt to initiate a customer service scam. The title jokingly admonished the scammer for being lazy.

4.1.3 Request for support. We coded 262 posts (17.2%) as a *request for support*, meaning that they asked for some type of support other than scam identification. In most cases, these posts related to specific victimization experiences, although a small number sought support with scam prevention. These posts were the majority of content we reviewed from `/r/Sextortion`, reflecting the community's primary purpose as a support group for people facing financial sextortion. We discuss the types of support that users requested in greater detail in Section 4.3.

4.1.4 Report of victimization. We coded 108 posts (7.1%) as *report of victimization*, meaning that they described users' experience with scam victimization without asking for support or asking to determine whether something was a scam. Reports most often described financial harm (56 occurrences, 51.9%), the sharing of intimate imagery with a scammer (23 occurrences, 21.3%), or the disclosure of personal information (14 occurrences, 13.0%).

Reports varied widely in detail, with some posts being extremely short. Several posts consisted of only titles that described an incident without going into further detail in the body or comments. Other posts were very detailed. For example, one user made a post to `/r/Scams` to describe their experience with a phishing scam.

They explained that they received a message with a short code purporting to be from their bank, reporting an unexpected debit card transaction. After interacting with the text message, they received a phone call from a spoofed phone number matching their bank and were socially engineered to send a transfer to the scammer. They concluded by discussing how they immediately realized that it was a scam, but were told by their bank that they were unlikely to get their money back. They expressed shame and the feeling of being stupid. These and similar posts provide valuable insight into victims' experiences by explaining the mechanics of the scam, identifying the features that caused the scam to seem plausible (i.e., authentic-looking message content, spoofed phone number, etc.), and describing the emotional and financial impact on the victim.

As with posts that were categorized as discussions of scam communications, users often expressed a desire that their story would help others avoid the same scam. For example, one user wrote a post identifying a specific Reddit user who scammed them on Craigslist. They concluded by expressing the hope that their post would stop others from being harmed. This type of prosocial motivation has been observed in previous studies examining victims' motivation for reporting fraud to the government [14].

Especially on /r/Sextortion, posts in this category often seemed to serve the purpose of 'venting' about victims' fears or negative emotions. For example, in a post made to /r/Sextortion, a user who was victimized four months prior discussed their difficulty in recovering emotionally. They shared that the experience had destroyed their self-esteem and given them a constant fear of making decisions. They described that being extorted drove them to begin smoking and drinking.

4.1.5 Other post categories. The remaining categories did not provide insight into users' experiences with scams and were not analyzed beyond initial categorization. Posts related to *scambaiting* were relatively rare in the communities we investigated (44 posts 2.9%). This result is unsurprising, as scambaiting content is not allowed under the rules of the largest subreddit we investigated, /r/Scams. Additionally, we coded 42 posts (2.8%) as *scam-related media*, consisting predominantly of links to news articles, blog posts, or YouTube videos warning users about particular scams. The *other* category (14 posts or 0.9%) was reserved for posts that were on-topic but did not fit into any of the other categories (e.g., general discussion of spam that was not related to a specific type of scam). Finally, we coded a small number of posts as *irrelevant* (35 posts or 2.3%), including generic spam and other off-topic posts.

4.2 RQ1: Types of scams discussed

In this subsection, we describe the types of scams discussed on Reddit. The most frequently reported scams—including sextortion, fraudulent ecommerce sellers, and corporate impostor schemes—align with both government data and Reddit's user demographics. Scenarios that did not meet our definition of a scam were rarely mentioned. Across the sample, most scams were mediated by the internet or related technologies.

4.2.1 Type coding results. Ecommerce seller scams, sextortion, and corporate impostor scams were the most common types of scams in our dataset. Table 2 shows the frequency of the 10 most common individual type codes. The high prevalence of *ecommerce seller* scams is unsurprising, as online shopping scams are the second most common type of fraud reported to the FTC, and the most common type of fraud reported to the FTC by those under the age of 39 [25] (i.e., the primary age group that uses Reddit [29]). Bitlab et al. previously leveraged Reddit discussion from /r/Scams to investigate fraudulent ecommerce scams [8]. The inclusion of /r/Sextortion in our dataset is mainly responsible for the high proportion of *sextortion*. We categorized only 26 posts (1.9%) that were posted to other subreddits as discussing sextortion.

Table 2: 10 most common scam type codes, excluding "unclear" and "likely not a scam." Percentages are calculated with respect to the 1,135 posts categorized as scam identification, discussion of scam communications, report of victimization, and request for support. Codes are not mutually exclusive; 352 or 31.0% of posts were assigned multiple type codes.

Scam Type	Type Category	#	%
Ecommerce seller	Commerce	145	12.8%
Sextortion	Threat	141	12.4%
Corporate impostor	Impostor	130	11.5%
Prize scam	Opportunity	102	9.0%
Employment scam	Opportunity	92	8.1%
Fake payment	Other	66	5.8%
Ecommerce buyer	Commerce	62	5.5%
Other threat	Threat	55	4.8%
Phishing	Impostor	53	4.7%
Investment scam	Opportunity	44	3.9%

Corporate impostor schemes also appeared frequently. 56.2% of the occurrences of the corporate impostor code came from posts categorized as discussion of scam communications, as compared to only 3.5% of posts about sextortion and 19.3% of posts about ecommerce seller scams. This may reflect that the users of these subreddits found these scams easier to identify and avoid, or that victims of these scams were less likely to seek help on Reddit. Impostor schemes are the most common type of fraud reported to the FTC, with older age cohorts more likely to report these schemes to the FTC than younger adults [25].

Many corporate impostor messages were likely attempts at *phishing*, although we only classified posts as such where the scam message clearly indicated that personal information was required (E.g., by explicitly asking users to log in). For example, in a post to /r/Scams, a user shared that they received a message saying that a delivery service had tried to drop off a parcel but was unsuccessful. They described that the message instructed them to click a link to initiate a new delivery. They went on to say that they found the scam particularly insidious because they were waiting for a package that morning. They acknowledged that it was likely a coincidence but that this experience helped them appreciate how people could fall for this type of scam. We did not categorize this as phishing because the author did not specify that sensitive information was requested.

Our scam-type classification is necessarily uncertain for posts where users have not been victimized by the scam they discuss. We observed many communications likely tied to scam attempts, but we lacked certainty. For example, 17 posts (1.5%) discussed so-called *wrong number* scams. These are spam texts seemingly addressed to someone else that are often sent with the hopes that the user will respond and a scammer will be able to form a relationship with the victim. While often used as an entree to investment scams [22], these texts may be legitimate. Moreover, even assuming they are fraudulent, it is impossible to tell what type of scam will follow based on the initial text alone, hence why we created a separate category for this ruse.

Many posts were also difficult or impossible to classify. 218 (19.2%) posts were classified as having an *unclear* scam type, meaning we could not unambiguously classify them within our coding scheme. The majority of these (49.5%) were in posts categorized as scam identification or discussion of scam communications (30.7%). For example, one post to /r/Scams from 2019 asked about a robocall in Chinese they were receiving on a weekly basis from various numbers. It seems probable that this robocall is tied to some sort of scam, such as the Chinese consolute impostor scam observed by Prasad et al. around the same time period [54]. However, it is impossible to categorize the incident with just the information provided.

4.2.2 Discussion of non-scams. Some posts clearly discussed incidents that did not meet our definition of a scam (see Section 2) 122 posts (10.7%) were categorized as *likely not a scam*, with most of these posts being categorized as scam identification (62.3%). Neither the posts coded as likely not a scam nor the associated comments were analyzed beyond post type and scam type categorization (e.g., they were not assigned codes for harm, communication methods, technologies used, etc.).

This set of posts was diverse, but we noted some common archetypes. Several posts involved cases in which users were concerned that legitimate interactions were scams. For example, in one post categorized as scam identification made to /r/Scams, a user who was seeking to purchase marriage records asked whether a local government website was legitimate, stating that they had no idea whether it was real and that they do not trust things.

Other posts in this category discussed business practices that users found distasteful but did not seem to constitute criminal behavior. For example, one user made a post complaining about an online shop that bought items at a discount from Lowes but sold them at the MSRP on Amazon. These posts reflect the colloquial usage of the term scam but are not of interest in understanding cybercrime.

Finally, some users discussed account compromise, identity theft, or other issues that involved no human interaction and were not clearly connected to scam victimization, like receiving mail addressed to other people. While it is possible that these experiences stemmed from scam victimization (e.g., phishing that results in account compromise), these posts lack detail on any social engineering that may have occurred.

4.2.3 Technologies used by scammers. The majority of scams discussed in the sample were mediated by the internet or related technologies. Only 31 posts (2.7%) were initiated via an in-person interaction or postal services. While a large proportion of posts (194 posts or 17.1%) did not clearly specify how the scam was initiated, the most common communications technologies used to initiate scams were social media/social media messaging (221 occurrences, 19.5%), instant messaging (148 occurrences, 13.0%), and email (122 occurrences, 10.7%).

Unsurprisingly, the type of communication technology used to initiate a scam differed based on scam type. The majority of ecommerce seller scams were initiated via ecommerce websites or services (62.1% of posts specifying an initial communication method). In contrast, and in accordance with prior work [11], sextortion was most often initiated by social media/social media messaging (53.1% of posts specifying an initial communication method). The majority of corporate impostor scams were initiated via instant messaging (35.8% of posts specifying an initial communication method) or email (30.0% of posts specifying an initial communication method).

The most common payment methods mentioned as being used in scams were digital payment services (83 posts or 7.3%) and crypto assets (61 or 5.4%). Gift cards were used only rarely in the sample (20 posts or 1.8%). The relative frequency of different payment methods is in line with the relative frequency of payment methods in FTC fraud reports [25]. Digital payment services like PayPal or Zelle are particularly popular with scammers, as they often lack protections for users who willingly initiate a transaction, even when they are deceived. For example, in its documentation on Fraud and Scams, Zelle states, “Even if you were tricked or persuaded into authorizing a payment for a good or service someone said they were going to provide... you may not be able to get your money back” [73]. Crypto assets are popular with scammers for similar reasons, as transactions are typically irreversible and pseudonymous; however, their use is more challenging for users, as evidenced by the fact that victims were often instructed on using crypto assets as part of scams.

4.3 RQ2: Type of support requested

In this subsection, we examine the nature of scam-related support-seeking on Reddit. Users sought help for a wide range of scams, with sextortion emerging as the most common in our sample. While many posts described financial losses, few explicitly asked for help recovering assets. Instead, users often requested more open-ended support. Notably, many users sought reassurance that the scammer could no longer harm them or other emotional support. These findings highlight the multifaceted nature of scam-related harm and the broad spectrum of support needs expressed by victims.

4.3.1 Types of scams in requests for support. Due to the nature of our sample, the plurality of posts categorized as requests for support discussed sextortion (103 posts or 39.3%). The remaining posts in this category were diverse, with no other single type of scam being asked about more than 25 times; scams involving ecommerce sellers (22 posts, 8.4%) and fake payments (19 posts, 7.3%) were the two next most common categories respectively. Moreover, 12 posts (4.6%) requested support unrelated to a specific scam and 20 posts (7.6%) requested support for things that were likely not scams.

Users typically requested support for scams perpetrated against themselves, but 30 posts (11.4%) requested support to help another person, like a family member or friend. Several of these posts related to older family members, like parents or grandparents, who were entrapped by a scam that the poster recognized but the victim did not. For example, one user made a post to /r/Scams describing how their father was being victimized by a scammer who had built up a relationship with him over Facebook Messenger before asking for increasing amounts of money. The poster had learned about the scam after the perpetrator requested tens of thousands of dollars for a fictional overseas investment. While the user's father hadn't sent the money, he was still in contact with the scammer. The poster requested help with preventing further harm, since confronting his father did not cause him to cut off contact. Difficulty in convincing victims they have been scammed has been noted by previous studies, especially with regard to romance scams [52, 71].

Users requesting support experienced a diverse set of harms. 84 of the posts categorized as requests for support (32.1%) mentioned some form of financial loss, 83 posts mentioned sharing of intimate imagery with a scammer (31.7%), and 42 posts (16.0%) mentioned the disclosure of personal information.

Financial loss is difficult to quantify, as users were often non-specific about how much money they lost, with 35 posts (41.7% of requests for support that mentioned a financial loss) not specifying the amount of money lost. Estimating based on conversion rates as of February 10, 2025, the median specified loss was $300, with values ranging from $4.99 to $175,000. This result is in line with the general distribution of fraud losses in FTC reports [25].

The most common types of information disclosed were generally not extremely sensitive. The most common items of information disclosed were a phone number (17 occurrences), name (12 occurrences), location (11 occurrences), and email address (9 occurrences). Only 6 posts mentioned disclosure of a Tax ID (e.g., social security number), 4 posts mentioned disclosure of bank account number, and 3 posts mentioned disclosure of one or more passwords. An additional 3 posts described account compromise without specifying what personal information had been disclosed. Our method likely under-counts the frequency of information leaks, as we only counted instances where users specifically discussed giving a scammer information. We did not infer that certain information was disclosed where this was not specified (e.g., many ecommerce seller scams likely involved disclosure of credit or debit card numbers).

Most victims of sextortion did not have their intimate imagery exposed. Only 15 or 14.6% of the posts classified as a request for support that discussed sextortion mentioned that their intimate image was posted publicly or sent to someone that they knew. This is likely an over-count, as scammers may show victims fake screenshots of intimate images being shared to convince them to pay. For example, one /r/Sextortion commenter reassured a poster by stating that they were shown screenshots of the scammer sending a close family member their intimate images, but the family member never received anything. Still, even assuming that all of these cases of reported exposure were genuine, this result is in line with prior findings that most extortionists do not follow through on threats to release intimate images [11].

4.3.2 Scam remediation requests. Table 3 shows the frequency of different types of support that were requested. Most of these requests can be classified as seeking advice on *scam remediation*, referring to the need for scam victims to return to their pre-victimization state. The most common request type code was *general scam remediation advice* (118 posts, 45.0%). This category was primarily composed of non-specific requests for advice. For example, in a post to /r/Scams, one user described how their friend never received payment after selling a concert ticket on social media. Their friend did research after the experience, finding a number of social media comments warning about the user. They asked if anything else could be done, expressing the hope that their post would serve as a warning if nothing could be done.

Table 3: The frequency and percentage of types of advice requested by Reddit posters dealing with scams. Codes are not mutually exclusive.

Support Sought	#	%
General scam remediation advice	118	45.0%
Reassurance about scam outlook	78	29.8%
Scam prevention advice	27	10.3%
Other users' experiences	27	10.3%
Understanding of victimization experience	22	8.0%
How to report scams	13	5.0%
Advice on financial recovery	13	5.0%
General emotional support	10	3.8%
Advice on collecting evidence	6	2.3%
Likely not a scam	20	7.6%

We also included in this category specific questions that did not match the other themes we developed but seemed to relate to recovering from scam victimization. For example, one user posting to /r/Sextortion asked how long they should wait before changing their social media account back to public visibility after being sextorted. As will be discussed in Section 4.4, changing privacy settings is a common strategy for mitigating the harm from sextortion.

Despite the high proportion of financial loss among requests for support, specifically asking about recovering lost funds (*advice on financial recovery*) was relatively rare, with only 13 posts. For example, one user asked about recovering funds from a Facebook Marketplace seller on /r/Scams. They were manipulated by lies about the scammer's ownership of an expensive car into paying with Apple Pay, and never received the item. They asked whether there was anything they could do to get their money back and referred to themselves as stupid for completing the transaction. As discussed is Section 4.2, recovery from peer-to-peer payment services not intended for commerce is, unfortunately, often impossible.

Some users also requested support with other aspects of remediation like *how to report scams* (13 occurrences) or *advice on collecting evidence* (6 occurrences). Typically, requests about reporting focused on service providers in order to have the scammer de-platformed. For example, one user posted to /r/Sextortion asking about whether it was possible to have the perpetrator's Instagram account banned. Posts asking about collecting evidence most often related to trying to find the scammer. For example, a user posted to /r/Scams sharing their experience with a scam involving fake concert tickets. They concluded their post by stating that they wanted to identify the perpetrator. Their motive is not stated, but presumably, they hope to recover the lost funds or bring the scammer to justice.

Many users also posted to Reddit to remediate the emotional or psychological harm resulting from scams. We categorized 78 posts that requested support as seeking *reassurance about scam outlook*, meaning that they wanted to verify that they were safe from further harm. This was especially common in the context of sextortion or other threat-based scams. In a typical example, a user made a post requesting reassurance after encountering a scammer on a dating app. The scammer pretended to be an underage girl and then called the poster pretending to be her parents. The poster acknowledged that the scenario seemed fake, but they still wanted confirmation that they were not in danger. Users who disclosed personal information also sought reassurance about how the information they provided may be used by a scammer. For example, in a post to /r/Scams, a user shared that their significant other gave a government impostor scammer part of his tax ID, their address, and the name of his bank after he was threatened with arrest. They asked what the scammer could do with the information they now have. Hasegawa et al. observed similar questions on Yahoo! Chiebukuro, especially from users who began entering information into phishing forms but stopped before final submission [31].

Exclusively on /r/Sextortion, users also requested more *general emotional support* that was connected to scam victimization, but was not specifically related to the outcome of a scam (10 occurrences). For example, one user recounted their story of sextortion, concluding by stating that they felt stupid for paying when it did not even make them feel more secure. They said they wanted to find other victims to talk with who could make them feel better. These requests illustrate both the deep psychological harm that scams can do and the value that users find in seeking support on Reddit.

These findings mirror Wei et al.'s investigation of image-based sexual abuse, as they found that many victims of financial sextortion use Reddit for therapeutic support [70]. We extend this insight to a broader range of scams, demonstrating that victims commonly seek reassurance about their safety and the potential consequences of their victimization. The fact that many users turn to Reddit for reassurance and emotional support suggests a gap in accessible, formal support systems for scam victims.

4.3.3 Other types of support sought. Some common types of support requested did not clearly fall into the category of remediation. For example, some users requested that others share their experience with scams. This was often implicitly or explicitly tied to a request for scam remediation advice or emotional support. For example, in a post to /r/Scams, a member of the military recounted their experience with a scammer whom they paid thousands of dollars after the scammer threatened to get them kicked out of the military. They requested help from anyone who had experienced something similar. We coded this question as both a request for general scam remediation advice and a request for *other users' experiences*.

In other cases, users' requests for similar experiences seemed more rooted in curiosity than a desire to draw on others' experiences for support. For example, a user shared a detailed account of their experience with a survey impersonating Verizon and purporting to give a prize. They concluded by saying that they were posting to prevent others from falling for the scam and to see if anyone else had been victimized by a similar scam. In total, there were 27 (10.2%) questions tagged as requests for other users' experiences.

Users also sought advice on protecting themselves from scams. 27 posts were categorized as seeking *scam prevention advice*. Most often, these requests were related to spam texts and calls that a user received. For example, a user made a post to /r/Scams requesting advice on how to stop spam calls that they were receiving on a daily basis. Other users requested advice more generally, not tied to specific communications. For example, one user made a post to /r/Scams asking how they could protect their loved ones from scams. Similarly, a user posted to /r/Scams requesting advice on how to avoid being defrauded when selling an automobile online.

Along the same lines of informational support, some scam victims requested clarification on how or why a scam occurred. We categorized these requests as seeking an *understanding of victimization experience* (22 occurrences). For example, one user made a post to /r/Scams asking why they were getting job scam emails to their main email account that claimed to have found them on a job board. Their account with the job board had a different email attached, so they could not understand how their main email account was (apparently) being found through the job board. They further asked for scam prevention advice in the form of questions about how to stop these scam emails.

Sometimes, the request for understanding of victimization was directly tied to the need to remediate harm. For example, one user sought help with account compromise stemming from a ransomware attack on /r/Scams. Despite taking a series of account remediation steps (i.e., changing their email password, ending all logged-in sessions, and disabling pop/imap), the attacker still had access to their email account. They asked how the scammer could access their email and how to block the scammer from accessing their email in the future. We categorized this set of questions as both seeking understanding of victimization experience and general scam remediation advice.

4.4 RQ3: Support type and quality

In this section, we describe the type of support received on Reddit. A plurality of the comments we reviewed were intended to explain features of the scam a user encountered or provide reassurance to a victim. The most common types of other advice related to remediating harm from scams, although users also gave more general security advice. Most advice was predominantly non-technical, emphasizing interpersonal, procedural, or institutional actions such as blocking the scammer, contacting a bank, or reporting the incident. In contrast, technical advice—such as changing account settings or deleting social media accounts—was relatively rare and typically surfaced only in specific contexts.

Beyond the content of specific advice, we identified several broader themes that characterized how scam support was delivered. These included the important role that bots played in providing explanations and directing victims to helpful resources, distinct patterns of interaction and emotional support between /r/Scams and /r/Sextortion, and the ways that scam reports benefited not only the original poster but also other internet users who discovered the threads later. Together, these findings suggest that Reddit serves as a community-driven infrastructure for scam identification, emotional support, and harm reduction.

4.4.1 Overview of support process. Receiving support was a dynamic process, with users asking clarification questions and posters providing additional details about their experience as appropriate. 475 of the 1, 883 comments we reviewed (25.2%) came from the original poster of the thread. The majority of these comments provided additional details about a user's experience (276 occurrences of *additional detail(s)*, 58.1% of comments from an original poster). The content of comments posted by other users varied widely, with a large proportion unrelated to our research questions (102 occurrences of the *other/irrelevant* code).

The most common type of support given to victims was to identify or otherwise explain the type of scam a user encountered. We assigned 416 comments the code *explain scam*. We frequently assigned this code to cases where users simply identified that something was a scam, but we also applied this code to more specific explanations of a user's understanding of the likely mechanics of the scam. For example, in response to a user of `/r/Scams` who inquired about how an account could have been compromised after they clicked on a phishing link but did not provide any information, one user speculated that the scammer could have installed a keylogger or stolen the Steam login cookie from their browser. It is unclear whether this explanation is correct, but it reflects the type of informational support often given.

The second most frequent type of support we observed was *reassurance* to victims (254 occurrences). We most frequently applied this code to comments that offered encouragement related to the likelihood of harm. For example, in response to a post requesting support for victimization by sextortion, one user offered reassurance by stating that it was actually against the fraudster's best interest for them to post their intimate imagery, as it places them at greater risk of being caught. They further shared an anecdote from a former scammer who claimed that scammers deleted data from hard drives to avoid being caught with evidence. We categorized this comment as both *explain scam* and *reassurance*, as the commenter infers how sextortion scams operate and uses these inferences to provide emotional support to the victim. We also applied the reassurance code where users offered more general emotional support, such as one commenter's statement that a victim of sextortion would be able to overcome their experience. The prevalence of reassurance is unsurprising in light of the many posts requesting reassurance or other emotional support. Wei et al. similarly noted a prevalence of therapeutic support among posts discussing IBSA, including financial sextortion [70].

The remaining types of support can be broadly classified as *scam remediation advice* and *general security/scam avoidance advice* based on the apparent purpose of the advice. The following sections discuss the most frequent types of advice in each context. A number of other specific pieces of advice appeared more than once but cannot be discussed due to space constraints (see the external Appendix). In addition to these codes, many comments included advice that was difficult to classify, especially where the advice was specific to a user's situation. We coded 105 comments as recommending *other remediation step(s)* and 85 comments as recommending *other security/scam avoidance advice.*

4.4.2 Scam remediation advice. As discussed in Section 4.3, users most often sought advice on remediating harms stemming from scams, so most comments were made in this context. The most frequent type of remediation advice given was the recommendation to *cut off communication* with the scammer by blocking them or otherwise ignoring their messages (198 occurrences). This advice was especially pertinent in the context of sextortion or other threat-based scams. For example, in response to a 16-year-old user who requested support after being extorted by a government impostor, a user advised them to cut off communication immediately, explaining that the police would not ask for money over the phone. This recommendation was sometimes given alongside the explicit recommendation not to pay the scammer any money (74 occurrences of *don't pay scammer*). For example, a user who sought support for sextortion against their boyfriend was advised that he should ignore them and not send any money. The commenter explained that the requests from the extortionist would only escalate and that paying them off permanently was impossible. This advice is generally well-founded and in line with recommendations from law enforcement [24].

Other common remediation advice was to *contact the financial institution/payment processor* associated with a scam (65 occurrences), or contact the provider of a service associated with a scam (55 occurrences of *contact service provider*). In the context of financial institutions/payment processors, this refers to things like disputing a transaction or locking a credit/debit card. These steps may allow a victim to recover lost funds or prevent the use of a compromised account. In the context of service providers, comments typically recommend that the victim report the scammer using the features built into services. This recommendation does not necessarily help the victim but does help platforms identify and remove scammers.

Recommendations to *contact law enforcement* (21 occurrences) or other government entities (10 occurrences) were less common. When reporting to law enforcement was discussed, it was occasionally described as pointless or ineffective. For example, one user made a post to `/r/Scams` asking how to file a sextortion complaint, stating that they wanted to harm the perpetrator legally. The first comments posted were dismissive of the idea, with one user stating that the scammer was likely untraceable due to using a stolen account. Another commenter jokingly compared sextortion to learning one has a terminal disease, stating that the poster needed to accept the situation and move on. While commenters are correct that bringing a scammer to justice is unlikely, reporting to law enforcement still provides a societal good by making authorities aware of the most pressing threats, which may lead to broader action against common scams.

Victims were also sometimes encouraged to *contact personal connection(s)* (33 occurrences), like family members or friends. This was often suggested in the context of sextortion to alert others to the risk of intimate imagery being released. For example, one commenter advised a victim of sextortion that they should tell their friends to block the scammer and ignore any messages. They suggested that the poster could claim the scammer used a deepfake app to generate the nudes to protect their privacy. Especially where users expressed serious emotional distress, it was also suggested that users may reach out to close others who can provide emotional support, including therapists. For example, one prior victim of sextortion responded to a new victim requesting emotional support by referencing their personal experience and encouraging them to talk with someone they feel safe with, such as a therapist or friend. Victims of sextortion may be reluctant to discuss their experience with others due to the stigma associated with victimization; however, it can be helpful to mitigate the harm from exposure and cope with the trauma.

Users were relatively infrequently provided with technical remediation advice. This result is unsurprising, as most schemes we observed were technically simple, involving internet technologies primarily used for communication. The most common form of advice was especially given to victims of sextortion, who were advised to *change security/privacy settings* (45 occurrences). This was often phrased vaguely, such as one user’s recommendation that another user should adopt the most stringent privacy settings available before logging back into social media. Users sometimes pointed to specific settings. For example, one commenter told a victim of sextortion that they should disable the ability for others to tag them in photos and turn off message requests on Instagram. Changing these settings would make it more difficult for the extortionist to harass the victim.

The next most common technical remediation advice was to change an identifier, such as phone numbers or social media usernames (44 occurrences of *change identifier*), or disable/delete one’s social media account (43 comments). These serve a similar purpose as changing privacy settings, making it more difficult for a scammer to contact a victim, although they may also harm a victim’s ability to interact with others.

4.4.3 General security/scam avoidance advice. Security and scam avoidance advice were most often provided in response to victims to help them avoid future harm. For example, users who lost money were often warned about the potential for re-victimization by “recovery” scams. (52 occurrences of *watch out for recovery services/scams*). These scams target crime victims by falsely claiming to help recover lost money or prevent the release of intimate imagery [26].

Recovery scams seem to be prevalent on Reddit, as some users responded to warnings about recovery scammers by sharing that they had been contacted by these scammers. For example, a user who requested support after their family member was victimized by an investment scam shared that they had already received a message from a supposed ethical hacker who promised to help get the lost money back. Subreddit moderation seemed to be effective at removing comments from scammers on public threads, with only 3 occurrences of comments advising users to contact recovery scammers. Still, this result demonstrates the risk of publicly seeking support for scam victimization.

Users were also commonly advised on how to avoid ecommerce scammers. This result is unsurprising, considering the large number of posts made related to ecommerce scams. We grouped the advice into two broad categories: advice on payments (36 occurrences of *advice on ecommerce payment*) and advice on selecting ecommerce websites (10 occurrences of *advice on where to shop*). With respect to payments, users were instructed to use different payment methods based on the context. Most commonly, users were told to avoid services like Zelle that lack fraud protection. As discussed in Section 4.3, these payment processors are often exploited by scammers, so steering people away from using them for purchases with strangers is appropriate.

Advice on where to shop was often vague, such as when one user recommended that people should only shop at well-reviewed websites. Users sometimes provided specific guidelines for identifying scam websites. For example, some users pointed to the age of domain registration as an indication of fraudulence. As fraudulent ecommerce websites are often short-lived [8], avoiding ecommerce websites with recently registered domains can be a useful heuristic to avoid scams.

Users were also advised not to click on links (6 occurrences of *don’t click links*) or otherwise interact with unsolicited communications (12 occurrences of *don’t answer communications from unknown sources*). For example, one minor who posted asking for support after receiving an unsolicited instant message containing pornographic content was told they should report the message and block the number that sent it. The same commenter also encouraged them not to click links from unexpected texts. This advice is commonly given to help users avoid phishing or other malicious websites [35].

Technical security best practices like multi-factor authentication (10 occurrences of *enable MFA*) or using unique passwords (6 occurrences of *use unique passwords*) were only rarely recommended. For example, a poster who requested advice after their family member fell for a customer support scam was told that their family member should use two-factor authentication and unique passwords as a precaution. Based on Redmiles et al.’s evaluation of security advice, these are considered by experts to be among the most important steps users can take to protect themselves from security and privacy harms [58]. In the aftermath of scams where personal information has been disclosed, these steps can help prevent account compromise.

4.4.4 Bots played an important role in the support process. Throughout the coding process, we noted that bots play an important role in providing support on `/r/Scams` and `/r/Sextortion`. In early 2018, the moderators of `/r/Scams` configured a tool called AutoModerator to provide explanations of common scam scenarios. For example, a user would type the command “!fakecheck” and the AutoModerator would provide a detailed explanation of the use of fake checks in scams: “...The fake check scam arises from many different situations... but the bottom line is... you receive a check... you deposit [a] check and see the money in your account, and then you use the funds to give money to the scammer... The bank will take the initial deposit back... and any money you sent to the scammer will come out of your own personal funds” *(verbatim).* Most often, this was used to explain the type of scam a user encountered in the context of the discussion of scam communications, scam identification, requests for support, or reports of victimization (i.e., comments coded as explain scam). The content of the explanations and the number of scenarios the bot could explain evolved over time. Some of the explanations provide advice on recovering from the scam. For example, an automated explanation of the advance fee scam from September 2021 stated, “...If you are involved in an advance-fee scam, you should attempt to dispute/chargeback any payments sent to the scammer... and you should ignore them if they attempt to contact you again" *(verbatim)*. Bots, and particularly AutoModerator, are widely used by communities across Reddit, although they are primarily used for administration tasks, such as screening posts for rule-breaking content [36, 43]. In this context, AutoModerator is being utilized as an informational tool that makes it simple for users to provide a standard explanation and basic advice for common schemes.

We also observed the use of AutoModerator to provide initial information to posters on /r/Scams and /r/Sextortion. Starting in early 2023 on /r/Scams and mid-2022 on /r/Sextortion, every post received an automated comment. On /r/Scams, this serves to warn users about the rules of the community and to encourage reporting of any scammers in the comments. This can be viewed as a form of "proactive moderation" that is designed to reduce rule-breaking content before it occurs [61]. On /r/Sextortion, these comments point users to a "new victim" guide that provides reassurance and resources to victims. In particular, it discusses the low likelihood of exposure with reference to a study of /r/Sextortion from the Canadian Centre for Child Protection [11], encourages users to seek professional mental health care if necessary, provides links to report sextortion to various law enforcement agencies and NGOs, and warns users about scammers that may attempt to re-victimize them. While not necessarily comprehensive, this post serves as a useful starting point for sextortion victims to understand and remediate their situation. In total, 139 comments on the posts we reviewed came from the AutoModerator. This number includes comments unrelated to support (e.g., messages informing a user why their comment has been removed).

4.4.5 Differences in support between /r/Sextortion and /r/Scams.

/r/Sextortion was generally more supportive than /r/Scams. While somewhat subjective, this difference can be illustrated through several coding results. /r/Sextortion users commonly offered to direct message (DM) users for emotional support. For example, one user responded to a sextortion victim who expressed suicidal ideation by encouraging the victim to message them and sharing that they also deal with depression, so they would not be judgmental. There were 25 occurrences of *you can dm me* on /r/Sextortion vs. 3 occurrences on posts from /r/Scams. Victims also seemed to be more likely to be chastised on /r/Scams, especially where they struggled to understand the deception they faced. In response to a victim of a fake payment scam who doubted that they had been scammed, a user wrote that they were foolish and called the scammers' ruses unbelievable. There were 24 comments assigned the code *chastise* on posts from /r/Scams vs. 1 comment on posts from /r/Sextortion. While we did not systematically analyze comments from scam identification posts, in the first round of coding, we also noticed that users were sometimes chastised for asking about scams that the commenter perceived as obvious. For example, one user made a post to /r/Scams asking about whether a sugar mama was real or a scammer. The framing of the post indicated that the poster suspected that it was fake, but multiple commenters chastised them for even entertaining the idea that anyone would pay them for a relationship.

These comparisons are qualitative and should be interpreted with caution. The number of annotated posts from /r/Sextortion is smaller than that from /r/Scams, so the observed differences may not generalize to all posts in either subreddit. Nonetheless, the thematic contrasts suggest differences in how each community responds to scam-related disclosures. Beyond the fact that victims of sextortion were more likely to request emotional support, one potential explanation for the difference between /r/Scams and /r/Sextortion is the increased prevalence of prior victims providing support. Many users on /r/Sextortion mentioned their prior experience with sextortion when offering support. For example, one sextortion user began a comment offering reassurance to a minor victim by sharing that they were the same age and were also currently dealing with an extortionist. In total, we coded 102 comments as referencing a *similar experience* on /r/Sextortion as compared to only 43 comments /r/Scams. This greater level of victim participation in the support process likely leads to more comments coming from users who understand the experience of victimization and are less likely to blame victims.

This result is similar to Marshall's observations of the treatment of some victims in romance scam discussion groups on Facebook [44]. While some victims were treated with pity (the victims who were "deserving" of sympathy), others were made fun of as lacking common sense or being ignorant (the victims who were "undeserving" of sympathy). Marshall tied this to gender differences, with men generally being treated worse than women, and the extent to which victims were perceived as behaving immorally (e.g., if the victim was cheating on their spouse). While we did not observe the same gender dynamics, the overly critical treatment of victims may undermine the value that these communities have as a support group.

4.4.6 Support given to Reddit posters helped other internet users.

Throughout all phases of the coding process, we noted several occurrences where a post about a scam led to discussion from other internet users who encountered the same scam, including requests for support not from the original poster (22 occurrences of *Non-OP request for support*). In response to a /r/Scams user reporting being defrauded by a Reddit user, three other commenters reported being scammed by the same person. Users were aware of this benefit of posting, with some posters deliberately adding the URL of a scam website or the name of an entity associated with a scam (e.g., username, company name, etc.) to a post to ensure that it was discoverable by other users. In a typical instance, a user requesting support with a crypto investment scam appended an edit to their post sharing the link to the scam website. They advised others not to click the link, but stated they wanted Google to recognize the URL as a scam. In recent years, Reddit has become increasingly prominent on Google [48], so it is unsurprising that the top results for scam websites include /r/Scams. In this way, Reddit serves as a de facto review website, protecting a wider range of users from harm.

5 Discussion

In this section, we summarize the results of our analysis and synthesize recommendations. In particular, we discuss the need for more automated tools to help users avoid scams, encourage future research based on Reddit scam discussions, and suggest greater law enforcement outreach on Reddit.

Through our analysis of four communities dedicated to scam discussion, we found that Reddit is used to help users identify scams, discuss the strategies used by scammers, and provide support to victims. The types of scams discussed on Reddit broadly reflect the types of scams most often reported to government entities, with a large proportion of posts focusing on ecommerce scams or corporate impostor scams (RQ1). Victims of scams often request general remediation advice, although many users also seek reassurance that they are safe from harm. /r/Sextortion especially serves as a support group for victims of financial sextortion (RQ2). Our analysis of Reddit comments suggests that these communities specialize in providing informational support and reassurance, with technical security advice rarely being needed or provided (RQ3).

Our results reveal challenges in moderating public support groups for scam victims. Prior work has noted many problems faced by volunteer moderators of online spaces, including an inability to keep up with the large volume of posts [36] and difficulties in protecting their communities from hate/harassment [64]. Uniquely, moderators of scam communities must ensure that victims are not re-victimized after posting about their experiences publicly. The public nature of these communities lowers the barrier to receiving support, but also means that scammers can easily find users who may be vulnerable to false promises of financial recovery. On /r/Sextortion and /r/Scams, users were protected through proactive warnings about recovery scams and apparently vigilant moderation of comment sections to remove scammers' messages. These protective strategies are especially relevant to the scam context, but may also be applicable to communities focused on victims of other forms of crime, such as survivors of intimate partner violence [50, 62].

Future work should build on these insights by interviewing or surveying moderators to better understand the practices involved in managing these sensitive spaces and whether there are opportunities for improvement. A content analysis focused on moderation-related comments could also provide greater insight into the moderation of scam-related spaces, although looking at publicly visible actions alone is likely insufficient to understand the full breadth of moderator labor [40]. While the largest communities we investigated (/r/Scams and /r/Sextortion) seem to already be excluding scammers from direct participation, a study focused on the moderation challenges could reveal opportunities for policy changes and new tooling in scam discussion spaces.

Our results also highlight the need for automated tools to help users reason about scams. The prevalence of posts seeking to identify potential scams reflects the difficulty many users have in distinguishing between legitimate and fraudulent communications online. While Reddit's scam communities serve as a useful resource, we note several drawbacks to relying on crowd-sourced scam advice. In addition to the risk of targeting by scammers, posters were sometimes chastised for their inability to recognize scams or blamed for being victims. While the best solution is to prevent users from encountering scams, tools that aid users in distinguishing between scams and benign messages could fill the gap left by current automated scam filtering techniques. Generative artificial intelligence (GAI) models present a promising path towards providing automated, context-specific advice on scams. Prior work has shown that GAI chatbots are effectively able to answer security-related questions [19] and categorize phishing messages [38, 60]. These results suggest that a tuned chatbot may be able to provide much of the same informational support without exposing users to harassment or the potential of re-victimization. Indeed, the use of AutoModerator on /r/Scams illustrates a basic form of automation in the support process. GAI tools could provide more specific, context-sensitive explanations than are presently possible with the template-based approach of /r/Scams.

Future work is necessary to explore whether GAI models can provide accurate advice when asked scam-related queries. Our results reveal the nuanced nature of scam-related help-seeking. Users often posed open-ended requests for support, rather than specific technical queries. Posters also needed emotional support alongside procedural advice about remediating scams. While AI chatbots are being increasingly adopted for mental health care solutions [30, 32], their effectiveness in responding to emotionally charged scam-related queries remains unclear. Those developing GAI-powered anti-scam interventions, including service providers such as Google [5], should consider our results in the design and evaluation of such tools. This includes assessing whether GAI-generated responses appropriately balance empathy and clarity, and testing how users respond to different explanation styles. Regardless of the quality of AI-generated support, online communities will remain important, as users may prefer receiving support from other humans.

We encourage future work based on Reddit scam discussion. Through their discussion, Reddit users have crowd-sourced a rich corpus of examples of scams that could be mined to create datasets of scam communications or measure patterns in scams over time. Our results indicate that Reddit discussion is especially useful for studying scams involving ecommerce sellers, sextortion, or corporate impostors. Posts categorized as discussion of scam communications are the most useful for this purpose, as they often include transcripts or images of scam communications. To our knowledge, only the study by Bitaab et al. has used Reddit discussion from /r/Scams for measurement purposes [8]. Prior work has used other social media platforms. For example, Nakano et al. developed CrowdCanary, a system to identify phishing URLs from English and Japanese reports posted on X (formerly known as Twitter) [46].

Measurement research of scam discussion on Reddit would not only be of academic interest, but also could give stakeholders, including the FTC, Anti-Phishing Working Group,[5] or Global Anti-Scam Alliance,[6] greater insight into emerging threats that are under reported via traditional means. Any research must consider the risks to users discussed in Section 3.3. Moreover, researchers should also consider sharing their results with the communities they measure or otherwise including community members in the research process [28]. Considering the use of the Canadian Centre for Child Protection's analysis of victim narratives[11] in its "New Victim's Guide," we believe that users will appreciate the insight into their experience that research provides.

We also suggest greater outreach from law enforcement agencies and other government entities to communities on

[5] https://apwg.org/

[6] https://www.gasa.org/

Reddit. While users were sometimes encouraged to report crimes—most notably in the `/r/Sextortion` "New Victim's Guide"—there was a pervasive attitude of futility regarding the benefits of reporting. Some users who reported scams on Reddit expressed a desire to protect others [14], so communications from stakeholders could acknowledge the difficulties in bringing scammers to justice while emphasizing the broader impact that reporting has. Communications could also clarify the most appropriate authorities to report to for internet crime, as confusion around jurisdictional issues can pose a barrier to reporting, especially in federal states like the United States or Australia [15, 16]. While such advice exists in other forms on the internet (e.g., [68]), it is valuable to meet users where they are. Within the United States, this outreach would most appropriately come from national-level entities like the FTC or FBI, which have the jurisdiction and resources to investigate internet fraud.

6 Conclusion

We presented a qualitative analysis of 1, 525 posts from four communities focused on discussion of scams on Reddit. We observed a variety of different types of discussion, with users especially using Reddit to discuss scammers' strategies, identify scams, and request support after victimization. Users discuss a wide variety of schemes, especially internet-mediated schemes involving ecommerce sellers, sextortion, or corporate impostors. Requests for help focus on requesting general remediation advice and reassurance about the likelihood of harm. Support given focuses on providing explanations of scams and offering the requested reassurance. We ultimately conclude that Reddit is a valuable resource for scam identification and remediation. Nevertheless, the prevalence of scam-related discussions highlights the ongoing need to better protect consumers online.

Acknowledgments

This material is based upon work supported by the Innovators Network Foundation and a Carnegie Mellon University CyLab Presidential Fellowship. We would like to thank Alexandra Nisenoff for helping collect the data underlying this paper. We would also like to thank Dr. Devesh Raval for providing feedback on an early draft of the manuscript.

References

[1] Jorij Abraham, Sam Rogers, Luka Koninng, Clement Njoki, and James Groening. Global state of scams report 2024. Technical report, Global Anti-Scam Alliance, 2024.

[2] Mutahar Ali, Arjun Arunasalam, and Habiba Farrukh. Understanding users' security and privacy concerns and attitudes towards conversational ai platforms. In *2025 IEEE Symposium on Security and Privacy (SP)*, pages 298–316, San Francisco, CA, USA, 2025. IEEE Computer Society.

[3] Keith B Anderson. To whom do victims of mass-market consumer fraud complain? *Available at SSRN 3852323*, 2021.

[4] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. The pushshift reddit dataset. In *Proceedings of the international AAAI conference on web and social media*, volume 14, pages 830–839, Atlanta, Georgia (virtual), 2020.

[5] Jasika Bawa. How we're using ai to combat the latest scams, 2025. Retrieved from https://web.archive.org/web/20250717024513/https://blog.google/technology/safety-security/how-were-using-ai-to-combat-the-latest-scams/, on September 22, 2025.

[6] Manon Berney, Jan Ondrus, and Adrian Holzer. Navigating the shadows of cyber vigilantism: A preliminary analysis of social dynamics and activities of scambaiting. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '24, New York, NY, USA, 2024. Association for Computing Machinery.

[7] 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.*, 6(CSCW2), November 2022.

[8] Marzieh Bitaab, Haehyun Cho, Adam Oest, Zhuoer Lyu, Wei Wang, Jorij Abraham, Ruoyu Wang, Tiffany Bao, Yan Shoshitaishvili, and Adam Doupé. Beyond phish: Toward detecting fraudulent e-commerce websites at scale. In *2023 IEEE Symposium on Security and Privacy (SP)*, pages 2566–2583, San Francisco, CA, USA, 2023.

[9] Casey Breen, Cormac Herley, and Elissa M. Redmiles. A large-scale measurement of cybercrime against individuals. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.

[10] Mark Button and Cassandra Cross. *Cyber frauds, scams and their victims*. Routledge, 2017.

[11] Canadian Centre for Child Protection. An analysis of financial sextortion victim posts published on r/sextortion, 2022. Retrieved from https://protectchildren.ca/en/resources-research/an-analysis-of-financial-sextortion-victim-posts-published-on-sextortion/ on September 22, 2025.

[12] Andrew Childs. 'i guess that's the price of decentralisation…': Understanding scam victimisation experiences in an online cryptocurrency community. *International Review of Victimology*, page 02697580231215840, 2024.

[13] Cassandra Cross. No laughing matter: Blaming the victim of online fraud. *International Review of Victimology*, 21(2):187–204, 2015.

[14] Cassandra Cross. Victims' motivations for reporting to the 'fraud justice network'. *Police Practice and Research*, 19(6):550–564, 2018.

[15] Cassandra Cross. 'oh we can't actually do anything about that': The problematic nature of jurisdiction for online fraud victims. *Criminology & Criminal Justice*, 20(3):358–375, 2020.

[16] Cassandra Cross, Kelly Richards, and Russell G Smith. The reporting experiences and support needs of victims of online fraud. *Trends and issues in crime and criminal justice*, (518):1–14, 2016.

[17] Marguerite DeLiema, Yiting Li, and Gary Mottola. Correlates of responding to and becoming victimized by fraud: Examining risk factors by scam type. *International Journal of Consumer Studies*, 47(3):1042–1059, 2023.

[18] Giuseppe Desolda, Lauren S Ferro, Andrea Marrella, Tiziana Catarci, and Maria Francesca Costabile. Human factors in phishing attacks: a systematic literature review. *ACM Computing Surveys (CSUR)*, 54(8):1–35, 2021.

[19] Lea Duesterwald, Ian Yang, and Norman Sadeh. Can a cybersecurity question answering assistant help change user behavior? an in situ study. In *Proceedings of the Symposium on Usable Security and Privacy (USEC) 2025*, San Diego, CA, USA, February 2025. The Internet Society.

[20] Marta Dynel and Andrew S. Ross. You don't fool me: On scams, scambaiting, deception, and epistemological ambiguity at r/scambait on reddit. *Social Media + Society*, 7(3):20563051211035698, 2021.

[21] Matthew Edwards, Claudia Peersman, and Awais Rashid. Scamming the scammers: Towards automatic detection of persuasion in advance fee frauds. In *Proceedings of the 26th International Conference on World Wide Web Companion*, WWW '17 Companion, page 1291–1299, Republic and Canton of Geneva, CHE, 2017. International World Wide Web Conferences Steering Committee.

[22] FBI El Paso. Fbi tech tuesday: Building a digital defense against "oops, wrong number!" texts, September 2023. Retrieved from https://web.archive.org/web/20241129093636/https://www.fbi.gov/contact-us/field-offices/elpaso/news/fbi-tech-tuesday-building-a-digital-defense-against-oops-wrong-number-texts on November 29th, 2024.

[23] FBI Internet Crime Complaint Center. Internet crime report, March 2023. Retrieved from https://www.ic3.gov/Media/PDF/AnnualReport/2023_IC3Report.pdf on September 22, 2025.

[24] Federal Bureau of Investigation. Sextortion: What kids and caregivers need to know. Retrieved from https://web.archive.org/web/20250303042405/https://www.fbi.gov/how-we-can-help-you/scams-and-safety/common-frauds-and-scams/sextortion/financially-motivated-sextortion on March 3rd, 2025.

[25] Federal Trade Commision. Consumer sentinel network data book 2023, Feb. 2024. Retrieved from https://www.ftc.gov/reports/consumer-sentinel-network-data-book-2023 on September 22, 2025.

[26] Federal Trade Commission. Refund and recovery scams. Retrieved from https://web.archive.org/web/20250208231804/https://consumer.ftc.gov/articles/refund-and-recovery-scams on February 8th, 2025.

[27] Casey Fiesler and Nicholas Proferes. "participant" perceptions of twitter research ethics. *Social Media+ Society*, 4(1):2056305118763366, 2018.

[28] 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.*, 8(GROUP), February 2024.

[29] Jeffrey Gottfried. Americans' social media use. Technical report, Pew Research Center, January 2024.

[30] MD Romael Haque and Sabirat Rubya. An overview of chatbot-based mobile mental health apps: insights from app description and user reviews. *JMIR mHealth and uHealth*, 11(1):e44838, 2023.

[31] Ayako A. Hasegawa, Naomi Yamashita, Tatsuya Mori, Daisuke Inoue, and Mitsuaki Akiyama. Understanding Non-Experts' security- and Privacy-Related questions on a Q&A site. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*, pages 39–56, Boston, MA, August 2022. USENIX Association.

[32] Michael V Heinz, Daniel M Mackin, Brianna M Trudeau, Sukanya Bhattacharya, Yinzhou Wang, Haley A Banta, Abi D Jewett, Abigail J Salzhauer, Tess Z Griffin, and Nicholas C Jacobson. Randomized trial of a generative ai chatbot for mental health treatment. *NEJM AI*, 2(4):AIoa2400802, 2025.

[33] Mo Houtti, Abhishek Roy, Venkata Narsi Reddy Gangula, and Ashley Marie Walker. A survey of scam exposure, victimization, types, vectors, and reporting in 12 countries. *arXiv preprint arXiv:2407.12896*, 2024.

[34] Temima Hrle, Mary Milad, Jingjie Li, and Daniel Woods. "just a tool, until you stab someone with it": Exploring reddit users' questions and advice on the legality of port scans. In *Proceedings of the 2024 European Symposium on Usable Security*, pages 322–336, 2024.

[35] Iulia Ion, Rob Reeder, and Sunny Consolvo. “...No one can hack my Mind”: Comparing expert and Non-Expert security practices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 327–346, Ottawa, July 2015. USENIX Association.

[36] Shagun Jhaver, Iris Birman, Eric Gilbert, and Amy Bruckman. Human-machine collaboration for content regulation: The case of reddit automoderator. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 26(5):1–35, 2019.

[37] Steven Kemp. Fraud reporting in catalonia in the internet era: Determinants and motives. *European Journal of Criminology*, 19(5):994–1015, 2022.

[38] Takashi Koide, Naoki Fukushi, Hiroki Nakano, and Daiki Chiba. Chatspamdetector: Leveraging large language models for effective phishing email detection. *arXiv preprint arXiv:2402.18093*, 2024.

[39] Klaus Krippendorff. Computing krippendorff's alpha-reliability. Technical report, Annenberg School for Communication, 2011.

[40] Hanlin Li, Brent Hecht, and Stevie Chancellor. All that's happening behind the scenes: Putting the spotlight on volunteer moderator labor in reddit. *Proceedings of the International AAAI Conference on Web and Social Media*, 16(1):584–595, May 2022.

[41] Jingjie Li, Kaiwen Sun, Brittany Skye Huff, Anna Marie Bierley, Younghyun Kim, Florian Schaub, and Kassem Fawaz. “it's up to the consumer to be smart”: Understanding the security and privacy attitudes of smart home users on reddit. In *2023 IEEE Symposium on Security and Privacy (SP)*, pages 2850–2866, 2023.

[42] Tianshi Li, Elizabeth Louie, Laura Dabbish, and Jason I Hong. How developers talk about personal data and what it means for user privacy: A case study of a developer forum on reddit. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3):1–28, 2021.

[43] Kiel Long, John Vines, Selina Sutton, Phillip Brooker, Tom Feltwell, Ben Kirman, Julie Barnett, and Shaun Lawson. "could you define that in bot terms"? requesting, creating and using bots on reddit. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 3488–3500, New York, NY, USA, 2017. Association for Computing Machinery.

[44] Jessica Marshall. The challenges posed by scammers to online support groups: The ‘deserving’and the ‘undeserving’victims of scams. *New perspectives on cybercrime*, pages 213–240, 2017.

[45] Gordon McDonald. Multi-label agreement, August 2023. https://gdmcdonald.github.io/multi-label-inter-rater-agreement/Multi-Label_Agreement.html.

[46] Hiroki Nakano, Daiki Chiba, Takashi Koide, Naoki Fukushi, Takeshi Yagi, Takeo Hariu, Katsunari Yoshioka, and Tsutomu Matsumoto. Understanding characteristics of phishing reports from experts and non-experts on twitter. *IEICE Transactions on Information and Systems*, E107.D(7):807 – 824, 2024.

[47] Gareth Norris, Alexandra Brookes, and David Dowell. The psychology of internet fraud victimisation: a systematic review. *J Police Crim Psych*, 34:231—-245, 2019.

[48] Katie Notopoulos. Reddit's traffic is way, way up. like, banoodles up. *Business Insider*, June 2024.

[49] Rajvardhan Oak and Zubair Shafiq. Victims, vigilantes, and advice givers: An analysis of scam-related discourse on reddit. In *Twenty-First Symposium on Usable Privacy and Security (SOUPS 2025)*, pages 57–71, Seattle, WA, August 2025. USENIX Association.

[50] Tully O'Neill. ‘today i speak’: Exploring how victim-survivors use reddit. *International Journal for Crime, Justice and Social Democracy*, 7(1):44–59, 2018.

[51] Masarah Paquet-Clouston, Matteo Romiti, Bernhard Haslhofer, and Thomas Charvat. Spams meet cryptocurrencies: Sextortion in the bitcoin ecosystem. In *Proceedings of the 1st ACM Conference on Advances in Financial Technologies*, AFT '19, page 76–88, New York, NY, USA, 2019. Association for Computing Machinery.

[52] 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*, 12(5):264, 2023.

[53] Nandita Pattnaik, Shujun Li, and Jason R.C. Nurse. Perspectives of non-expert users on cyber security and privacy: An analysis of online discussions on twitter. *Computers & Security*, 125:103008, 2023.

[54] Sathvik Prasad, Elijah Bouma-Sims, Athishay Kiran Mylappan, and Bradley Reaves. Who's calling? characterizing robocalls through audio and metadata analysis. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 397–414. USENIX Association, August 2020.

[55] Nicholas Proferes, Naiyan Jones, Sarah Gilbert, Casey Fiesler, and Michael Zimmer. Studying reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media+ Society*, 7(2):20563051211019004, 2021.

[56] Devesh Raval. Which communities complain to policymakers? evidence from consumer sentinel. *Economic Inquiry*, 58(4):1628–1642, 2020.

[57] Joseph Reagle. Disguising reddit sources and the efficacy of ethical research. *Ethics and Information Technology*, 24(3):41, 2022.

[58] Elissa M. Redmiles, Noel Warford, Amritha Jayanti, Aravind Koneru, Sean Kross, Miraida Morales, Rock Stevens, and Michelle L. Mazurek. A comprehensive quality evaluation of security and privacy advice on the web. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 89–108. USENIX Association, August 2020.

[59] Andrew S Ross and Lorenzo Logi. ‘hello, this is martha’: Interaction dynamics of live scambaiting on twitch. *Convergence*, 27(6):1789–1810, 2021.

[60] Muhammad Salman, Muhammad Ikram, Nardine Basta, and Mohamed Ali Kaafar. Spallm-guard: Pairing sms spam detection using open-source and commercial llms. *arXiv preprint arXiv:2501.04985*, 2025.

[61] Charlotte Schluger, Jonathan P Chang, Cristian Danescu-Niculescu-Mizil, and Karen Levy. Proactive moderation of online discussions: Existing practices and the potential for algorithmic support. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):1–27, 2022.

[62] Nicolas Schrading, Cecilia Ovesdotter Alm, Ray Ptucha, and Christopher Homan. An analysis of domestic abuse discourse on Reddit. In Lluís Màrquez, Chris Callison-Burch, and Jian Su, editors, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2577–2583, Lisbon, Portugal, September 2015. Association for Computational Linguistics.

[63] Xinyue Shen, Zeyuan Chen, Michael Backes, Yun Shen, and Yang Zhang. "do anything now": Characterizing and evaluating in-the-wild jailbreak prompts on large language models. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, CCS '24, page 1671–1685, New York, NY, USA, 2024. Association for Computing Machinery.

[64] Madiha Tabassum, Alana Mackey, and Ada Lerner. 'custodian of online communities': How moderator mutual support in communities help fight hate and harassment online. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*, pages 297–314, Philadelphia, PA, August 2024. USENIX Association.

[65] Mohammad Tahaei, Tianshi Li, and Kami Vaniea. Understanding privacy-related advice on stack overflow. *Proceedings on Privacy Enhancing Technologies*, 2022.

[66] Mohammad Tahaei, Kami Vaniea, and Naomi Saphra. Understanding privacy-related questions on stack overflow. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–14, New York, NY, USA, 2020. Association for Computing Machinery.

[67] Lauri Tuovinen and Juha Röning. Baits and beatings: Vigilante justice in virtual communities. *Proceedings of CEPE*, pages 397–405, 2007.

[68] USAGov. Learn where to report a scam. Retrieved from https://www.usa.gov/where-report-scams on September 22, 2025.

[69] Fangzhou Wang. Breaking the silence: Examining process of cyber sextortion and victims' coping strategies. *International Review of Victimology*, 31(1):91–116, 2025.

[70] 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 *33rd USENIX Security Symposium (USENIX Security 24)*, pages 4391–4408, Philadelphia, PA, August 2024. USENIX Association.

[71] Monica T Whitty and Tom Buchanan. The online dating romance scam: The psychological impact on victims – both financial and non-financial. *Criminology & Criminal Justice*, 16(2):176–194, 2016.

[72] Xin-Li Yang, David Lo, Xin Xia, Zhi-Yuan Wan, and Jian-Ling Sun. What security questions do developers ask? a large-scale study of stack overflow posts. *Journal of Computer Science and Technology*, 31:910–924, 2016.

[73] Zelle. Fraud & scams overview. Retrieved from https://web.archive.org/web/20250201174250/https://www.zellepay.com/safety-education/fraud-scams-overview on Feburary 1st, 2025.

[74] Sijie Zhuo, Robert Biddle, Yun Sing Koh, Danielle Lottridge, and Giovanni Russello. Sok: Human-centered phishing susceptibility. *ACM Trans. Priv. Secur.*, 26(3), apr 2023.