A First Look into Fake Profiles on Social Media through the Lens of Victim's Experiences

Arezou Behfar
Utah State University
Logan, Utah, USA
abehfar@usu.edu

Ankit Shrestha
Utah State University
Logan, Utah, USA
ankit.shrestha@usu.edu

Mahdi Nasrullah Al-Ameen
Utah State University
Logan, Utah, USA
mahdi.al-ameen@usu.edu

ABSTRACT

With the rise in online social media usage, fake user profiles are becoming prevalent concerns. A body of work suggested machine learning-based models to identify fake profiles, however, the accuracy and applicability of those techniques are still at large. A little study to date, considered end users in the loop, to understand the strategies of intruders, and the reactions of victims. We begin to address this gap in our work, where we aim to look through the lens of users' perceptions who had prior experiences of interacting with fake profiles on social media (i.e., victims of fake profiles). To this end, we conducted semi-structured interviews with 26 participants. Our findings unpack the viewpoints of fake-profile victims, with regard to intruders' traits and strategies to befriend them on social media, as well as their reactions when they realize that they have interacted with a fake profile. Based on our findings, we provide recommendations on empowering and supporting social media users in order to alleviate their vulnerability to online exploitation.

CCS CONCEPTS

• **Security and privacy**; • **Human-centered computing → Human computer interaction (HCI)**;

KEYWORDS

Online social media, fake profiles, semi-structured interview

ACM Reference Format:
Arezou Behfar, Ankit Shrestha, and Mahdi Nasrullah Al-Ameen. 2024. A First Look into Fake Profiles on Social Media through the Lens of Victim's Experiences. In *Companion of the 2024 Computer-Supported Cooperative Work and Social Computing (CSCW Companion '24), November 9–13, 2024, San Jose, Costa Rica.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3678884.3681889

1 INTRODUCTION

Over the past few years, we have witnessed a rapid growth in the popularity of online social media platforms. As of April 2020, the global number of social media users reached an astounding 3.81 billion [19], where Facebook, Instagram, Twitter, YouTube, and LinkedIn are among the leading social media platforms that have garnered a massive user base worldwide [19]. According to a study conducted by Fire et al. [11], an estimated 8.7% of Facebook accounts do not belong to real users. This observation suggests a significant presence of fake accounts on online social media. The social networking sites typically do not verify the authenticity of a user when they create an account, and do not check whether a single user has created multiple accounts. This unrestricted access lets fraudulent users to create fake profiles and exploit the platform for personal or collective gains [19]. This issue is further exacerbated by the fact that user profiles on social media platforms, such as Facebook, contain a large amount of personal data, posing privacy and security risks for legitimate users [27, 29].

Fake profiles refer to the creation of social media accounts using fabricated or stolen credentials, with the intention of engaging in activities that are detrimental to the users and communities within social networks [2]. Fake profiles manifest in various forms and serve different purposes, where the well-known social media sites, including Facebook, Instagram, and Twitter have become increasingly plagued by the presence of fake profiles, rendering them unreliable and untrustworthy for many of legitimate users [19, 31]. Thus, the protection from fake profile is of utmost importance to protect users from potential harm.

Human decision-making typically takes a more comprehensive approach than an automated system [12]. Thus, understanding the experiences of those affected by fake profiles is crucial as it can provide insights into the human aspects of social media privacy and security that an automated fake profile identification system may overlook. Further, automated identification of fake profiles through Machine Learning (ML)-based techniques poses a significant challenge [14, 15]. To this end, a body of work [20, 30] started to focus on understanding how users could detect fake profiles. However, a little study to date, looked into the experiences of victims who have been targeted by fake profiles. As we begin to address this gap, we investigated following research questions through the lens of victims' experiences:

RQ1. What are the strategies of fake profiles (also termed as intruders, and scammers in this paper) in deceiving online social media users (i.e., victims)?

RQ2. How do victims respond and handle the situation when they realize that they have interacted with a fake profile?

To address these research questions, we conducted semi-structured interviews with 26 participants who have been the victims of fake profiles. Our participants' interaction with fake profiles unfolded the motives and manipulation techniques of intruders (RQ1; see §4.1), including impersonation, video communication, geographic proximity, pre-payment, and intimate conversation – exploited to deceive the victims and earn their trust. Our study shed light on victim's reactions (RQ2; see §4.2), where we discuss our

CSCW Companion '24, November 9–13, 2024, San Jose, Costa Rica
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1114-5/24/11.
https://doi.org/10.1145/3678884.3681889

findings in realm of the forensics of fake profiles (see §4.2.1), managing discoverability in online social media (see §4.2.2), and seeking help and justice (see §4.2.3). The findings from our study lead to the recommendations for service providers on empowering and supporting social media users (see §5) to make their platform more robust against fake profiles.

2 RELATED WORK

Online Social Networks are inundated with fake profiles aimed at harvesting user information and compromising their security [16]. Awan et al. [2] described fake profile as a social media account created using fabricated or stolen credentials where the primary intent behind account creation is to engage in activities that can be harmful to a specific, or a large group of users. According to Bharti et.al [4], users who intentionally present themselves as someone else on social media are referred to as fake users, and the accounts associated with them are known as fake profiles; the authors pointed out three types of fake profiles: Human (a profile created and maintained by a human), Bot (computer-generated fake profile, administered by a program that performs repetitive tasks to achieve specific objectives), and Cyborg (a hybrid of Human and Bots, representing a profile created by a human but is subsequently operated by bots).

Roy et.al [19] reported, a fake profile is typically created by extracting data from existing profiles on the network, including profile name, profile photo, age, gender, and other readily available information. These studies [4, 14, 16, 19] pointed out a wide range of reasons behind the creation of fake profiles, like spreading misinformation for political and business gain, celebrities or politicians seeking to inflate their fan base, business organizations trying to manipulate feedback and ratings, and gaming applications supported by Facebook encouraging users to create fake accounts in pursuit of incentives tied to bringing in more players.

Online social networks employ detection measures to counter the growing threat posed by fake profiles, where the majority of such techniques are based on Machine Learning (ML) algorithms [15]. The rapidly evolving social media platforms lead intruders to adapt their tactics to deceive users, rendering ML-based approaches less effective [14]. This is especially true when intruders steal people's identities or exploit real profiles that have already established a level of trust within their network, making detection and removal by service providers more challenging [31]. Further, the differences in feature characteristics across online social networks complicate the collection and extraction of features, posing challenges to develop effective algorithm for fake account detection [14].

The study of Habib et al. [12] reviewed new-age ML algorithms for fake profile detection and suggested that, despite advancements in AI models' speed, scalability, and pattern recognition, human expertise and experience remain valuable as they can apply common sense and logic, which AI often lacks, particularly in assessing content plausibility. To this end, a body of work [20, 30] started to focus on understanding how users could detect fake profiles. The study of Sandy et al. [20] examined how verbal and nonverbal cues affect the credibility judgment of fake Twitter profiles, where 24 participants evaluated 16 fake Twitter profiles in a lab setting. The study [20] revealed that content of a profile was the primary factor influencing the credibility assessments by participants. Wang et al. [30] compiled datasets comprising confirmed fake profiles, legitimate users, and suspicious profiles; they conducted a user study with participants categorized as experts and crowd-sourced. Participants were tasked with assessing the authenticity of profiles based on basic information like name, gender, profile picture, and friend count, where experts performed better than the crowd-sourced participants. In a separate study focusing on social media scams [8], Chergarova et al. pointed to the need of vigilance and user education to protect them from intruders.

Our Work. A little study to date, looked into the experiences of victims who have been targeted by fake profiles. We addressed this gap in our work, where we leveraged the lens of victim's experiences in understanding the strategies employed by intruders to deceive online social media users, and how the victims react when they realize that they have interacted with a fake profile.

3 METHODS

We conducted semi-structured interviews with 26 participants. The study was approved by the Institutional Review Board (IRB) at our university.

3.1 Participants Recruitment

To take part in our study, participants had to be at least 18 years old, with prior experience of interacting with fake profiles on social media – specified in our recruitment message. We reached out to participants through social media posts, including on Facebook, Instagram, and LinkedIn. We also posted flyers on public notice boards at our university. In addition, we used snowball sampling, where we recruited a few participants based on the recommendation of participants who already took part in our study.

We recruited a total of 26 participants (16 female, 10 male). The age of our participants varied between 18 and 60, where most of them were in the age range between 26 and 40. Their level of education ranged from high school diploma to graduate degree, with the majority holding a graduate degree. Participants' experiences in using online social media platforms varied between 4 and 15 years, with an average of 7 years.

3.2 Procedure

As participants showed interest to take part in our study, we emailed them the Informed Consent Document (ICD). As they agreed to ICD, we scheduled a time for online interview over Zoom. During the interview (audio-recorded), participants (victims) shared their experience of interacting with a fake profile (intruder) on social media, including the techniques employed by intruder to earn victim's trust which was then exploited for malicious intent. We asked participants about the consequences they faced as a result of interacting with the fake profile. Participants also reflected on the actions taken as they identified the profile as fake, as well as their experiences and expectations in getting help from the service provider for online social media. At the end, participants responded to a set of demographic questionnaire, hosted on Qualtrics. On average, each session took between 30 and 60 minutes to complete.

3.3 Analysis

We transcribed the audio recordings from interviews and performed thematic analysis [3, 6, 7, 23, 24] on our transcripts. We took an inductive approach for our analysis. In this ground-up approach, codes are derived from the data without preconceived notions, which allows the narrative to emerge from the raw data itself without trying to fit it into the preconceptions [3, 6, 7]. Each participant's data was coded by two independent researchers, where the inter-coder reliability was 86%. They developed codes from the transcripts of the first few participants, compared those codes, and then iterated again with more participants' data until a consistent codebook was developed. Once the codebook was finalized, two researchers independently coded the remaining participants' data. After all participants' data had been coded, both researchers discussed and resolved the discrepancies in their codes. Finally, we organized and taxonomized our codes into higher-level categories.

4 FINDINGS

In this section, we report the findings from our study. Following the guideline from prior research [22, 25, 28], for consistency, we used the following terms based on the frequency of themes in participants' responses: a few (0-10%), several (10-25%), some (25-40%), about half (40-60%), most (60-80%), and almost all (80-100%).

4.1 Intruder's Strategies (RQ1)

Our participants' interaction with fake profiles unfolded unique stories, shaped by the motives and manipulation techniques of intruders. While impersonation is a common technique used by intruders (see §4.1.1), our findings also unpack their strategies to gain victim's trust, which include video communication, geographic proximity, pre-payment, and intimate conversation (see §4.1.2).

4.1.1 Impersonation. Impersonating someone else and assuming their identity is a common trait that our participants associated with the fake profile. Our findings reveal the multifaceted nature of impersonation in creating a fake profile, which include assuming the identity of a celebrity, or a friend in the online social network. One of our participants (P15) referred to the advent in AI technology that he believes, it is used to make the interaction with a fake profile seem realistic in order to deceive the victim: *"It is...a robot. So fake accounts would be a program trying to imitate a single person interacting with others."* Several participants shared their experiences about how they were manipulated to trust a fake profile, where the intruder established a false sense of connection by assuming the identity of a familiar person.

4.1.1.1 Impersonating Celebrities. One of our participants, P8 reported how a fake Instagram profile mimicked the persona of a celebrity and scammed her while she was searching for an online job: *"Their strategy was to use the photo and name of a trusted person among the people...this person has a special social status where everyone trusts him. The intruder was deceiving people through the trust that everyone has in this showman."* She further detailed, intruder published an advertisement on Instagram profile about an online job, and she reached out to inquire about this opportunity. Then the intruder convinced her to deposit money as a proof of her commitment to the job; she put her trust and sent money considering the public image of that profile. However, the promise of employment, as well as that profile eventually turned out to be fake.

4.1.1.2 Impersonating Friends. The participant, P10 shared an incident where the intruder assumed her identity and used her photos to perpetrate a scam within her Instagram friend list. The intruder messaged her acquaintances over social media, claiming to be in need of financial assistance. Our participant mentioned, one of her friends trusted the claim and believing it to be her, sent money to the scammer.

4.1.2 Gaining Trust. Our findings point out the strategies employed by intruders to gain a victim's trust, which include video communication, prepayment, exploiting geographic proximity, and intimate conversation.

4.1.2.1 Video Communication. A few of our participants pointed out video communication as the means of scammers to establish trust, which they found convincing due to the belief that a fake profile would typically prefer to protect their anonymity. The participant, P22 shared his experience while he came across a rental advertisement on Facebook as he was looking for a house to rent; here, the scammer used the advertisement as a bait, and later convinced him to deposit a fraction of rent in advance by showcasing a compelling video of the purported property: *"I told her [scammer] that I want to see this house before I sign any contract or pay any money. She told me, I will arrange an appointment for you to come and see the house tomorrow. She texted me the next morning, saying that I have an urgent trip but I can send you a video from home right now...She sent me a video that showed different parts of the house...So I said to myself as this person sent me a video...probably everything is OK. I paid by PayPal, and she sent me the receipt. Then she told me to pay the rest of the money for the first month's [rent]...I told her I would not do this until I see the house, and she blocked me [on Facebook] the next day."*.

4.1.2.2 Prepayment. In another instance, the scammer posed as a potential buyer of a laptop and paid a small fraction of its price in advance, to convince our participant (P13) about her legitimacy; P13 elaborated how she was deceived, *"Someone bought a laptop from me [on Facebook]...and that person pretended to be an elderly woman. And after I sent the laptop to them, they did not pay for it, and I figured out that the profile was fake, but I could not do anything...This is how I was, unfortunately, deceived."*.

4.1.2.3 Geographic Proximity. Several participants shared similar experiences where they put trust on a social media profile as the location of residence noted in that profile was around where they live. Such trust led them to pay for products which they never received; P4 reported, *"I looked at their page [on Instagram]; their clothes were suitable, and the prices were reasonable. Everything was okay. Because it [intruder's residence] was in our area, I didn't think it was fake at all. After I made the deposit, they blocked me. They used my location to reach out to me, and it seemed they had chosen the user based on their area. It was a trap."*

4.1.2.4 Intimate Conversation. We also found instances where the intruder adopted a warm and amiable approach to establish a rapport and create a sense of intimacy. One of our participants (P14) shared her life event while she was going through a phase of loneliness; during that time, the online conversation with someone she did not meet before gave her comfort: *"It was like he was giving very personal and friendly messages. He started in the morning, for example, 'Good morning. Did you eat breakfast?' and then continued the conversation...the person was so lovely..."* Our participant further mentioned that the dialogue of that profile was seducing, sometimes aimed at convincing her to invest money on his cryptocurrency company, which she later discovered to be fake.

4.2 Victim's Reactions (RQ2)

In addition to financial loss (see §4.1), about half of our participants reported experiencing psychological issues, including stress, anger, and fear as a result of their interaction with fake profiles. In certain cases, these interactions led victims to lose their trust on social media platforms (see §4.2.3 for details). One of our participants (P20) referred to his difficulty in trusting people, which further exaggerated upon being deceived by a fake profile in social media: *"I did end up in therapy later and had to work through some trust issues...I have struggled with depression...I had mistrusted people. And so as soon as that [interaction with fake profile] happened, it left me traumatized, where I was afraid to speak to people online. And I had more or less grown to think everybody was lying to me. So it wasn't until years later working with a therapist, I've been able to overcome some of that finally..."*

Our participants reported that encountering fake profiles heightened their sense of caution in online navigation, where they attempted to understand the features of a fake profile based on their experiences (see §4.2.1), took actions to better manage their discoverability in social media (see §4.2.2), and sought justice from the service provider to protect legitimate users from intruders hidden behind the curtain of fake profiles (see §4.2.3).

4.2.1 Forensics of Fake Profiles. Based on the experience of interacting with fake profiles, our participants identified several attributes that they had learnt as the possible indicator of a fake profile; P4 reflected on her learning from a fake profile she encountered: *"A profile picture of a flower and plants or landscape instead of their own photo...and also no posts or following and followers...it is apparent that they did not want to take time and they wanted to create an account quickly. I think this is a fake profile."*

4.2.1.1 Profile Picture and Name. Several participants mentioned, the fake profile they interacted with had a low-quality profile picture, likely due to taking a screenshot from other's profile in process of impersonation. The participant, P6 shared her experience of being deceived by an intruder who impersonated her friend and used a screenshot of that friend's photo as a profile picture: *"The profile picture was a screenshot because the picture quality was very low and blurry...the photo was cropped, some parts of the photo were missing."* Some participants also reported that the fake profile they interacted with did not have a traditional name, where P9 said, *"They [fake profiles] do not have a specific name, for example, [it named] 'brave boy'."*

4.2.1.2 Social Media Connections and Activities. Our participants shared a common viewpoint that a fake profile has a very few social media connections, where P12 said, *"...[fake profiles] have very few followers and followings...they typically have less than 10 for both following and followers."* Upon identifying a profile as fake, about half of our participants perceived that they should have considered the absence of recent posts and reactions from friends (e.g., likes, comments) as the indicator of a suspicious profile, where P13 said, *"when I came back and looked at that [fake] profile, I saw that her last photos were from several years ago."* Similarly, P22 stated, *"From 2014 to 2022, she had only one photo every year, but there were no comments or likes on her photos."* While an intruder is possibly aware of such red flag, they might post fake comments from their other profiles; one of our participants (P4) recalled such unusual pattern in comments on a fake profile she interacted with: *"...I remember that one user wrote 3 to 4 comments including words one by one after each other,like great, great, and good, in separate lines that raised their number of comments."*

4.2.1.3 Communication Pattern. After their encounter with fake profiles, about half of our participants started to pay more attention to communication patterns, including discrepancies in facts or events shared by a profile, and too-good-to-be-true stories about their lavish life style. Participants emphasized the details in biography section when discerning the authenticity of an account, where several of them mentioned, multiple profiles with alike attributes or sharing the similar story could be a potential red flag; P14 came across such profiles in social media: *"A very handsome and classy bodybuilder, 80% of his profile was similar to the previous one. I doubted his photo and the similarity of his story to the previous one."*

4.2.2 Managing Discoverability. The features and information on social media platforms, including demographics and location sharing, public pages, friend-lists, and friend-suggestions grant intruders a wide range of options to find their target. Our participants reflected on their belief about how they were targeted by intruders, which dictated the change in their navigation strategies after encountering a fake profile, to protect their discoverability in online social media. Our participants emphasized the significance of privacy settings, where a common practice that emerged from interacting with a fake profile was shifting from public to private setting for their account. P9 mentioned changing the setting that prevents others from finding her Facebook account through Google search; she further added, *"...this is a preventative method that decreases the risk of being targeted by fake accounts."*

4.2.2.1 Demographics and Location Sharing. Some of our participants reported that demographic information including gender and age revealed through a profile picture and information in 'about me' section on social media made them vulnerable to be a target of intruder, where P19 shared his view: *"...it could be because of my age, I wonder if they saw that I was a young person, so I might be more susceptible..."*. Another participant (P12) commented, *"Based on my experience and feedback from my close friends of both genders, these fake profiles usually target women more than men."* A few participants also referred to location sharing, which could have exposed their economic status and so on, made them vulnerable to be deceived by a fake profile seeking financial gain.

To this end, we found instances where participants mentioned taking several protective measures upon interacting with a fake profile, such as reducing the amount of shared information, and being cautious about location sharing. A few participants also changed their profile picture so that their gender identity is not revealed publicly. For instance, P7 mentioned, *"My posts and photos changed exactly after that [interaction with fake profile]. It made me more cautious than the previous version of myself..."*

4.2.2.2 Following Public Pages. Some of our participants perceive that they were targeted by intruders because of their exposure on public pages. For instance, P11 suspects that her appearance in the follower list of an online shop page on Instagram made her discoverable by the scammer. Several participants believe that they were targeted by a fake profile due to their activity on public pages, such as posting a comment or liking a post that were publicly visible. For instance, P9 speculates that the scammer found her through the comments she posted on the public page of a sports club.

Upon encountering a fake profile, participants reported being more careful in following public pages, where they have reduced the number of pages they follow. A few participants also use referrals to confirm authenticity before following a public page, where P11 mentioned, *"Channels [pages] that I choose to follow have to be suggested by others who checked that profiles before, then I will follow them. I try to follow other people and pages more carefully. And I don't follow anyone without prior knowledge because one of the ways that fake profiles find you is through the channels you follow, they find you on those pages."*

4.2.2.3 Exposure through Friends in Social Media. Several participants perceive that the intruder found them in the publicly visible friend-list of their social media friends or through Facebook's 'people you may know' feature that suggests users to add as friends. To reduce the scope of public visibility through social media connections, our participants curated their friend-list after interacting with a fake profile. A few of them also asked their friends who tagged them in photos or posts to remove those tags. However, participants reported that despite the steps taken, they may still be discovered by intruders through their social media connections due to their lack of control on the visibility of their friends' friend-lists, as well as whom the social media would suggest through 'people you may know' feature. To this end, a few participants temporarily deactivated their accounts, believing that their inactivity would reduce the scope of public exposure.

4.2.2.4 Beyond Managing Discoverability: Familiarity Check and Cross-referencing. In addition to taking measures to protect discoverability as discussed above, some of our participants mentioned that upon encountering the fake profile, they now consider familiarity, particularly having mutual friends as a deciding factor before befriending someone on social media. A few of them also reaches out to a mutual friend to confirm the identity and trustworthiness of a new profile who has sent a friend request. One of our participants (P23) who now follows this strategy to protect from fake profiles, mentioned, *"The first thing is to see if we have a mutual friend or not; then, If I am relatively close to that mutual friend, I will ask who this person is"* A few participants also cross-reference through Google search to verify the authenticity of a profile, where P1 reported, *"I save their [someone who sends a friend request] photo and then check that photo in Google Image. I found many fake accounts in this way."*

4.2.3 Seeking Help and Justice. About half of our participants blocked and reported the fake profile they interacted with. However, they were worried that the intruder could find them over other social media platforms or impersonate another identity to deceive them in the future. We found instances where participants aimed for legal actions or sought help from the service provider with a hope of getting more effective and sustainable solution to safeguard themselves from further intrusion.

4.2.3.1 Contacting Legal Authority. Several participants tried to pursue legal action and seek justice for their financial loss resulting from the interaction with fake profile, however, did not receive adequate help from the authority. For instance, P4 reported, *"I went to the Cyber police and gave them their [intruder] page's address. The police officer said because they [intruder] use VPN, they are not identifiable."* In some cases, victims refrain from taking legal actions due to the fear of greater harm that could be caused by an intruder. For instance, P13 who did not receive payment for the laptop she sold (see §4.1.2 for details), shared her experiences: *"I was terrified, especially when I sent the package [laptop] by post...my phone number and house address, all were mentioned under sender's information. I was afraid for a long time, and many people around me made me afraid of them by telling me: 'These people are gangs. So you should not follow up too much because they are dangerous'. That's why, I was afraid to leave home alone for a while. I felt very insecure."*

4.2.3.2 Contacting Social Media Service Provider. One of our participants (P20) reported, *"They [service providers] don't tell you how to avoid it [fake profile]. They don't give you examples. They don't warn you of things to look for. They, they really don't try to help you, from my experience."* While the lack of guidance and warning from service providers to protect from fake profiles caused frustration among participants, it was further exaggerated when they did not get the help from them after experiencing financial loss and psychological issues upon interacting with a fake profile. P20 further elaborated, *"I spent hours just trying to get somebody to respond to me, let alone help me. And then they [service providers] never helped. Felt like they ignored their customer...There was no help."*

In another instance, P10 whose identity was stolen and used for impersonation by a fake profile (see §4.1.1 for details), shared her experience as she sought help from the service provider: *"I can talk about my Instagram experience... they [service providers] say, 'if something happens to you, we will remove the fake account'...I emailed them...very clearly, they told me, 'we cannot do anything'. I could not understand from their email who the scammer is, is it me or fake account? The email was very complicated."*

5 DISCUSSION

In this section, we discuss the implications of our findings, which lead to the recommendations on empowering and supporting users, and so on, make social media platforms more robust against fake profiles.

5.1 Empowering Users

It is crucial for social media users to be aware of fake profiles and their strategies in order to protect themselves. To this end, social media platforms can incorporate a 'protection from fake profiles' section, linked to security and privacy setting sections, as well as presented to users as a prompt nudging them to be aware of fake profiles, review and reflect on their recent interactions over social media, and take caution before committing a financial transaction with social media contacts. While reporting suspicious profile is already an available option, users should be encouraged to share their story behind such reporting with the service provider – it has multiple implications in protecting against fake profiles: First, it can be used to update 'protection from fake profiles' section, as intruders could come up with new strategies to deceive users [9]; Second, some of those stories can be shared with social media users (keeping the victim's identity anonymous) to raise public awareness, while prior studies [10, 18, 32] indicate the efficacy of social stories in making users aware of security and privacy issues.

Our findings indicate that the publicly visible friend-list can be exploited by an intruder to target a victim, and eventually more people from their mutual friends. To this end, social media platforms can prompt users about privacy-preserving setting on the visibility of their friend list, so that someone outside of their network have limited or no access to their friend list. Here, limited access may represent viewing the name of a friend in the list but not being able to access that profile or send a friend request. Further, users can be given provision to set their privacy so that they will be in control whether their name will be publicly visible in the friend-list of their social media friends. Through adopting these measures, social media platforms can empower users with the knowledge and tools necessary to recognize and combat fake profiles, fostering a safer online environment.

We found that having mutual friends is considered to be a credible indicator of the authenticity of a profile when the user receives a friend request, which is in line with the findings from prior studies [5, 11] that users tend to accept friend requests from a stranger when they have mutual friends. In this context, our participants reported concern that the existing features of social media platforms, e.g., 'People you may know' on Facebook, can expose them to an intruder, further increasing the risk of being targeted by a fake profile. To this end, social media platform can present users with a 'verify before you make a friend' prompt that will encourage the user to send a message to some of their mutual friends for verifying the authenticity of that profile, as well as to be informed of any red flag (e.g., bullying, inappropriate behavior, etc.) they should know about before accepting that friend request from an unknown individual. Such a message can also be sent to a 'prior' mutual friend, who have already unfriended that profile.

5.2 Supporting Users

We noticed that the metrics, including number of friends, followers and their reactions (e.g., like, comments) to the posts are considered by our participants when assessing the legitimacy of a profile. Prior studies [1, 26] pointed out 'fake engagement' referring to the fact that social media platforms are often used for business promotion where there are companies that sell followers (e.g., 50,000 followers for a price of $250). That means, fake profiles can buy such 'fake engagement' to deceive users and gain their trust. Thus, social media platforms should support users to protect from fake profiles by checking the authenticity when a new user or page signs up. Identity verification during registration can be a potential measure to consider in this regard, where service providers can inaugurate their own identity verification system or leverage the existing services (e.g., ID.me, etc.). We acknowledge that introducing such scheme would make the registration process lengthy and more complicated than it is now; we encourage future studies to examine this tension between usability during registration and user expectations to protect from fake profiles in social media. Further, future studies should explore automated technique to identify a sudden surge in the number of followers, which may indicate a possible purchase of 'fake engagement' – adoption of such approach by social media platforms along with keeping new profiles under scrutiny for a certain period of time will be contributing towards keeping users safe from fake profiles.

Our findings point to the dissatisfaction of participants with the support provided by social media platform as they reached out to them upon enduring financial loss and psychological trauma caused by fake profiles. Such dissatisfaction resulted from delay in response, deviation from commitment, and lack of empathy for the victims. To address these issues, social media service providers should come forward to establish a notion of trust for their users, starting with offering a provision for real-time conversation, e.g., through online chat with trained customer representatives. The training of customer representatives should cover a wide range of spectrum, including the strategies of intruders, possible reactions of victims, and how to reassure and support them in a timely manner along with taking appropriate actions towards the intruder. Enabling users to engage in real-time conversations with the trained representative would substantially enhance user experience and restore their trust on social media platforms in combating fake profiles.

6 LIMITATIONS AND CONCLUSION

We interviewed 26 participants in our study, where we followed the widely-used methods for qualitative research [3, 6, 7], focusing in depth on a small number of participants and continuing the interviews until no new themes emerged. We acknowledge the limitations of such study that a different set of samples might yield varying results. Thus, we do not draw any quantitative, generalizable conclusion from this study. Further, most of our participants were well-educated. Since users' security and privacy perceptions are positively influenced by their knowledge and technical efficacy [13, 17, 21], we speculate that the privacy perceptions and behavior of users reported in this paper represent an upper bound in the context of online social media navigation.

A few of our participants were recruited via snowball sampling. In snowball sampling, participants who have taken part in the study nominate people for recruitment whom they know well, and thus, it may suffer from sampling bias. In addition, self-reported data might have limitations, like recall and observer bias.

Despite these limitations, our findings provide valuable insights to the understanding of intruder's strategies in deceiving online social media users, and how the victims react when they realize that they have interacted with a fake profile. Based on our findings, we offer recommendations on empowering and supporting users, to make online social media platforms more robust against fake profiles. We encourage future research to extend the findings from our study through involving participants from diverse demographics, including older adults, and users with limited literacy.

ACKNOWLEDGMENTS

We would like to thank the reviewers for their valuable feedback. This material is based upon work supported by the National Science Foundation under Award No. CNS-1949699.

REFERENCES

[1] Fatih Cagatay Akyon and M Esat Kalfaoglu. 2019. Instagram fake and automated account detection. In *2019 Innovations in intelligent systems and applications conference (ASYU)*. IEEE, 1–7.

[2] Mazhar Javed Awan, Muhammad Asad Khan, Zain Khalid Ansari, Awais Yasin, and Hafiz Muhammad Faisal Shehzad. 2022. Fake profile recognition using big data analytics in social media platforms. *International Journal of Computer Applications in Technology* 68, 3 (2022), 215–222.

[3] Kathy Baxter, Catherine Courage, and Kelly Caine. 2015. *Understanding your users: a practical guide to user research methods*. Morgan Kaufmann.

[4] Nasib Singh Gill Bharti and Preeti Gulia. 2023. Exploring machine learning techniques for fake profile detection in online social networks. *International Journal of Electrical and Computer Engineering (IJECE)* 13, 3 (2023), 2962–2971.

[5] Yazan Boshmaf, Ildar Muslukhov, Konstantin Beznosov, and Matei Ripeanu. 2011. The socialbot network: when bots socialize for fame and money. In *Proceedings of the 27th annual computer security applications conference*. 93–102.

[6] Richard E Boyatzis. 1998. *Transforming qualitative information: Thematic analysis and code development*. sage.

[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[8] Vasilka Chergarova, Vinicius Arcanjo, Mel Tomeo, Jeronimo Bezerra, Luis Marin Vera, and Anthony Uloa. 2022. Cryptocurrency fraud: A study on the characteristics of criminals who are using fake profiles on a social media platform to persuade individuals to invest into cryptocurrency. *Issues in Information Systems* 23, 3 (2022).

[9] Emmanuel Etuh, Francis S Bakpo, et al. 2022. Social Media Networks Attacks and their Preventive Mechanisms: A Review. *arXiv preprint arXiv:2201.03330* (2022).

[10] Matthias Fassl, Alexander Ponticello, Adrian Dabrowski, and Katharina Krombholz. 2023. Investigating Security Folklore: A Case Study on the Tor over VPN Phenomenon. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–26.

[11] Michael Fire, Dima Kagan, Aviad Elyashar, and Yuval Elovici. 2014. Friend or foe? Fake profile identification in online social networks. *Social Network Analysis and Mining* 4, 1 (2014), 1–23.

[12] AKM Rubaiyat Reza Habib, Edidiong Elijah Akpan, Bhaskar Ghosh, and Indira Kalyan Dutta. 2024. Techniques to detect fake profiles on social media using the new age algorithms-A Survey. In *2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 0329–0335.

[13] Iulia Ion, Rob Reeder, and Sunny Consolvo. 2015. "...No One Can Hack My Mind": Comparing Expert and Non-Expert Security Practices. In *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security* (Ottawa, Canada) *(SOUPS '15)*. USENIX Association, USA, 327–346.

[14] Shruti Joshi, Himanshi Gupta Nagariya, Neha Dhanotiya, and Sarika Jain. 2020. Identifying fake profile in online social network: An overview and survey. In *Machine Learning, Image Processing, Network Security and Data Sciences: Second International Conference, MIND 2020, Silchar, India, July 30-31, 2020, Proceedings, Part I 2*. Springer, 17–28.

[15] Sarah Khaled, Neamat El-Tazi, and Hoda MO Mokhtar. 2018. Detecting fake accounts on social media. In *2018 IEEE international conference on big data (big data)*. IEEE, 3672–3681.

[16] Priyanka Kumari and Nemi Chandra Rathore. 2018. Fake profile identification on Facebook through SocialMedia app. In *Information and Communication Technology for Sustainable Development: Proceedings of ICT4SD 2016, Volume 2*. Springer, 83–90.

[17] Michelle L. Mazurek, Saranga Komanduri, Timothy Vidas, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Patrick Gage Kelley, Richard Shay, and Blase Ur. 2013. Measuring Password Guessability for an Entire University. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security* (Berlin, Germany) *(CCS '13)*. Association for Computing Machinery, New York, NY, USA, 173–186. https://doi.org/10.1145/2508859.2516726

[18] Emilee Rader, Rick Wash, and Brandon Brooks. 2012. Stories as informal lessons about security. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*. 1–17.

[19] Pradeep Kumar Roy and Shivam Chahar. 2020. Fake profile detection on social networking websites: a comprehensive review. *IEEE Transactions on Artificial Intelligence* 1, 3 (2020), 271–285.

[20] Christopher Sandy, Patrice Rusconi, and Shujun Li. 2017. Can humans detect the authenticity of social media accounts?. In *3rd IEEE International Conference on Cybernetics (CYBCONF-2017)*.

[21] Sovantharith Seng, Huzeyfe Kocabas, Mahdi Nasrullah Al-Ameen, and Matthew Wright. 2019. Poster: Understanding User's Decision to Interact with Potential Phishing Posts on Facebook Using a Vignette Study. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (London, United Kingdom) *(CCS '19)*. Association for Computing Machinery, New York, NY, USA, 2617–2619. https://doi.org/10.1145/3319535.3363270

[22] Ankit Shrestha, Arezou Behfar, and Mahdi Nasrullah Al-Ameen. 2024. "It is Luring You to Click on the Link With False Advertising"-Mental Models of Clickbait and Its Impact on User's Perceptions and Behavior Towards Clickbait Warnings. *International Journal of Human–Computer Interaction* (2024), 1–19.

[23] A Shrestha, A Flood, B Hackler, A Behfar, and MN Al-Ameen. 2024. TOWARDS THE DESIGN AND EVALUATION OF CLICKBAIT EDUCATION CONTENT: LEVERAGING USER MENTAL MODELS AND LEARNING SCIENCE PRINCIPLES. In *INTED2024 Proceedings*. IATED, 1794–1804.

[24] Ankit Shrestha, Audrey Flood, Saniat Sohrawardi, Matthew Wright, and Mahdi Nasrullah Al-Ameen. 2024. A First Look into Targeted Clickbait and its Countermeasures: The Power of Storytelling. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–23.

[25] Ankit Shrestha, Danielle M Graham, Prakriti Dumaru, Rizu Paudel, Kristin A Searle, and Mahdi Nasrullah Al-Ameen. 2022. Understanding the Behavior, Challenges, and Privacy Risks in Digital Technology Use by Nursing Professionals. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–22.

[26] Sharifa Sultana, Pratyasha Saha, Shaid Hasan, SM Raihanul Alam, Rokeya Akter, Md Mirajul Islam, Raihan Islam Arnob, AKM Najmul Islam, Mahdi Nasrullah Al-Ameen, and Syed Ishtiaque Ahmed. 2022. Imagined Online Communities: Communionship, Sovereignty, and Inclusiveness in Facebook Groups. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–29.

[27] Catalina Laura Toma. 2014. Counting on friends: Cues to perceived trustworthiness in Facebook profiles. In *Eighth International AAAI Conference on Weblogs and Social Media*.

[28] Leonie van Grootel, Lakshmi Balachandran Nair, Irene Klugkist, and Floryt van Wesel. 2020. Quantitizing findings from qualitative studies for integration in mixed methods reviewing. *Research synthesis methods* 11, 3 (2020), 413–425.

[29] Putra Wanda and Huang Jin Jie. 2020. DeepProfile: Finding fake profile in online social network using dynamic CNN. *Journal of Information Security and Applications* 52 (2020), 102465.

[30] Gang Wang, Manish Mohanlal, Christo Wilson, Xiao Wang, Miriam Metzger, Haitao Zheng, and Ben Y Zhao. 2012. Social turing tests: Crowdsourcing sybil detection. *arXiv preprint arXiv:1205.3856* (2012).

[31] Mudasir Ahmad Wani, Suraiya Jabin, Ghulam Yazdani, and Nehaluddin Ahmadd. 2018. Sneak into devil's colony-A study of fake profiles in online social networks and the cyber law. *arXiv preprint arXiv:1803.08810* (2018).

[32] Rick Wash. 2010. Folk models of home computer security. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*. 1–16.