

Evaluating the Social Media Safety and Privacy of Vtubers - Creating a List of Guidelines for Vtubers to Protect their Safety on Social Media

Binette, Joseph

School of Computing and Information Science
University of Maine
Orono, Maine, United States of America
joseph.binette@maine.edu

Corlett, Caleb

School of Computing and Information Science
University of Maine
Orono, Maine, United States of America
caleb.corlett@maine.edu

Abstract—This paper will detail some of the risks that VTubers face on social media as well as some behaviors VTubers and livestreamers can do to improve their individual privacy and information security on social media. VTubers are a relatively new phenomenon in the online space, which have exploded in popularity worldwide since their inception in 2016. They are performers that have moved into the online entertainment landscape, commonly filling the roles of video and livestream content creators. What makes them unique is their usage of face, voice, and body tracking technology to give them a sense of pseudonymity by taking on the appearance of a 2D or 3D character. With this pseudonymity comes unique advantages and challenges associated with how they maintain their privacy, security, and safety. VTubers rely on social media and have high propensities for cybercrime. If they do not have strong individual privacy and information security practices in place, they risk bringing harm to themselves and anyone associated with them. This paper consists of a literature review about vtubers, their unique online privacy standards, as well as relevant academic literature about leading-edge information security practices on social media. Then, a list of behavioral guidelines, specialized to VTubers, that can be followed to improve their individual privacy, security, and safety on social media sites. This paper has the possibility to result in improvements to VTubers privacy on social media and could help normalize safer social media behaviors.

Index Terms—VTuber, Livestreamer, Twitch, Youtube, Privacy, Security, Safety

I. INTRODUCTION

A VTuber is a content creator that performs over the internet behind a digital avatar, often resembling that of a character similar to those from Japanese anime that mimics their facial expressions and some of their actions. [1] Compared to traditional influencers and content creators, this digital avatar allows VTubers much more privacy, as they can perform without showing their actual face. This unique level of privacy often attracts scrutiny into the VTuber's identity. When using social media, VTubers must take exceptional care to protect their individual privacy and to avoid putting themselves at risk from malicious actors that wish to see the metaphorical man behind the curtain. VTubers, particularly female VTubers, can sometimes experience increased harassment due to parasocial

relationships that viewers may form. Turner (2023) notes that a significant portion of her participants encountered said harassment, and it was a source of discomfort for them. [1] Even though the digital avatar VTubers use may be complex and require a degree of digital literacy to utilize, VTubers come from all walks of life and may have backgrounds other than tech or they just might not have a good grasp on individual privacy or information security enough to protect themselves from malicious actors on the internet. This study aims to use publicly available records of past incidents, current practices within VTuber communities, and the most current research on individual privacy and information security best practices to produce a specialized and easy to understand list of behavioral guidelines for VTubers to follow to improve their individual privacy and information security on social media sites.

A. Problem Statements This Study Aims to Address

- What privacy requirements do VTubers require?
- What risks to their individual privacy and information security do vtubers face?
- What are some vtuber specific individual privacy and information security challenges that vtubers face on sns?
- What are some behaviors a vtuber can adopt to improve their individual privacy and information security on sns?

B. Risks of Being a VTuber

Clipping culture is a common practice within VTuber communities where a short segment from a livestream or piece of recorded content is uploaded online, showing off a funny or important moment from a longer piece of a VTuber's content in an easily consumable format that is bingeable and gratifying to watch. Clipping culture allows for VTubers to gain flash-virality (for possibly positive or negative reasons) and with it, attention from a lot of new people who aren't a part of their trusted audience, which may include attention from potential malicious-actors. If a VTuber has a weak grasp on individual security, they could risk the security and safety of themselves, their children, their family, their uninvolved friends, or other VTubers.

C. Why VTubers?

The reason for choosing VTubers as the target of this study is because of the unique circumstances regarding VTubers and their online privacy. Vtubers, like more traditional influencers and content creators, put on a display of themselves and their personalities in front of audiences that can reach into the thousands. The digital avatar that VTubers use, however, allow them to perform under an untraditional amount of individual privacy. With this increased privacy comes increased scrutiny from their viewing audience, which may result in a breach of a VTuber's individual privacy caused by malicious actors if a VTuber does not have a solid grasp of individual privacy and information security practices.

D. Significance of a List of Behavioral Guidelines

The significance of this study is that it will result in an easily understandable and distributable way to inform VTuber and other high privacy, low tech literacy individuals of techniques and behaviors they could use to improve their information security and protect their identities better within the realm of social media. The decision to specialize these guidelines was due to the reason that individual security guidelines are much more effective and widely utilized when they go beyond generic cases [4].

E. Study Approach

This study will consist of an observational analysis on publicly available information regarding individual cases of breaches of VTubers' privacy and security, an observational analysis on publicly available information regarding widely-used vtubing practices used to protect individual privacy and information security and the reasons behind them, as well as a literature review of the most recent academic articles regarding VTubers, VTuber culture, as well as best practices regarding individual privacy and information security on social media.

F. The Rest of the Paper

The rest of the paper is structured as follows:

- In Section "Guideline Creation", we will discuss in further detail the process of how we created our behavioral guidelines.
- In Section "Results and Findings", we will briefly summarize the findings from our observational analysis and highlight some of the most important guidelines from our list that would offer the biggest improvements to a vtuber's individual privacy and information security on social media.
- In Section "Discussion", we will discuss challenges and shortcomings of our observational analysis as well as offer suggestions for improvement for anyone else looking to develop a similar list of behavioral guidelines.

II. LITERATURE REVIEW

A. VTubers, Identity, and How It Relates to Privacy

Despite the relatively short amount of time in which the concept of VTubing has existed, there is a rather remarkable amount of literature regarding VTubers. This literature

spans all kinds of perspectives through which the concept of VTubers can be examined, both from a technological, and psychological standpoint. The relationship between a VTuber's pseudonymity and digital privacy was loosely connected at best in many cases, however. Most closely is the multiple articles that examine how VTubers express and conceal their identity through their models. Byron (2022) and Turner (2022) [1], [2] both separately address this issue. Turner took a direct approach in interviewing 10 self-described VTubers from video streaming platform Twitch. She concluded that VTubing is a way for one to be as they want to be (whether it be their authentic self, an exaggerated self, or a character) without any of the discomfort of revealing their face to the internet. Her approach was good because it allowed for an honest dialogue where the interviewees could share whatever information they would like. Its single largest shortcoming, however, is the sample size of 10, which was cut down from 100 due to general lack of response. Byron's article on the other hand is observational. Much of his research into VTuber identity was focused on the members of Hololive's English team, which he gained insight on simply by watching their streams. Byron addresses how VTubers form their identity in the context of more strict characterization as is mandated for Hololive cast members. He notes that they tend to use techniques such as mystification to make real-world references whilst still staying in character. He also discusses the idea of "forbidden knowledge," that being information that is known about the Hololive casts' real identities. Said forbidden knowledge is commonly banned in Hololive-adjacent communities. Byron's research has the advantage of being contextualized within the sphere of VTuber management, however, it lacks the directness of Turner's research, and is largely more speculative. These articles both contain information about how VTubers protect and project their identities. Identity can be an important aspect of privacy, and while the articles do occasionally bring up more direct privacy issues (such as the above forbidden knowledge), they lack detail about how VTubers can further protect their privacy. In comparison, this article is being made to concretely establish how VTubers can protect their privacy, and that includes aspects of their identity.

B. Virality and Harassment

Clipping culture allows for vtubers to achieve short term virality, a sharp influx of attention from users outside of the usual audience. The attention gained from a viral clip can be positive or negative in nature, with interaction from bigger accounts producing interactions that mirror the sentiment of the bigger account [3]. There are three attributes that are a likely indicator of coming harassment if a viral piece of media is associated with, how these attributes are associated could be as the topic of the viral clip, or as attributes of the featured vtuber. These three attributes are Minorities; Feminism, or any mentality not bound by older, more established values; and opinions and discussion regarding male-dominated fields like politics or sports [3]. If a vtuber is the target of harassment, social media platforms often offer ways for an individual to

protect themselves, including blocking, muting, and reporting [3]. Another thing Tonami et al. address is possible actions an individual can try if they represent a company and are unable to employ platforms' self-protection measures due to their position and encourages individuals to reach out to their company and discuss possible actions that the company can take to assist the individual or any compensation that can possibly be provided for any legal fees or other relevant costs [3].

C. Organizations' attempts to prevent phishing of employees

In their article, "The dark side of social networking sites: Understanding phishing risks", Silic et al. do an experiment testing different ways they could make employees of a fortune 500 company fall for phishing scams, providing employees of the companies with phishing links and recording who was most likely to click on scam links and enter their information in different contexts and situations. "Females were more likely to become victims (54.9%) and overall, the younger population (aged 20-30) represents the most vulnerable category" [4], some individuals opened the phishing link but did not enter any information due to the webpage not "looking genuine" [4], and it was also found that employees from the call center were significantly more vulnerable to falling prey to phishing attacks than from any other company department [4]. Along with the experiment, Silic et al. held interviews with employees that participated in the experiment and found that "employees have little knowledge of their company's existing policies regarding SNS use" [4], "employees trust SNSs and are generally not aware of the potential security risks" [4], and "security training regarding SNSs is not effective" saying that security training is "generic" and "not-effective" [4]. Silic et al. also interviewed Chief Information Security Officers from 11 different companies and asked what they thought companies could do to improve the information security of their employees on SNS. The Chief Information Security Officers generally believed that "organizations do not seem to have any mechanisms in place to control online security threats" [4], seven of them believed that their company did not have adequate information security policies [4], all of them believed that "SNS is a major security hole" for companies [4], and that "educational awareness and training are the best strategies, and essential to leveraging information security." [4].

D. The link between individual's behavior and their chances of becoming the victim of cybercrime

Saridakis et al. in their paper, "Individual information security, user behaviour and cyber victimisation: An empirical study of social networking users", use a theory from traditional crime theory called the Routine Activity Theory to explain how individuals come to be victims of cybercrime [5]. According to the Routine Activity Theory, crimes happen when a motivated offender and an accessible target meet without the presence of guardianship [5]. Saridakis et al. had over 700 individuals belonging to certain clusters within

various SNS sites and online communication spaces fill out a survey regarding if they've been victims of cybercrime in the past. From the survey, it was found that high usage of SNS sites modeled around the sharing of information, like LinkedIn or Twitter, resulted in higher probability of becoming a victim of cybercrime due to higher usage correlating with more information being shared [5], the perceived control over one's information was effective in lowering the probability of becoming a victim [5], technical efficacy did not have an effect on the probability of becoming a victim due to higher confidence leading to riskier behavior [5], and individual with higher propensities for taking risks online had a positive correlation with victimization rates [5].

E. Building knowledge about SNS information security from a parallel entertainment industry

In the academic article, "'I feel physically safe but not politically safe': Understanding the Digital Threats and Safety Practices of Only Fans Creators", Soneji et al. interviewed 43 sexual content creators from Onlyfans about the risks they face and the practices sexual content creators can use to protect themselves from said risks [9]. From the interviews conducted, the biggest risks sexual content creators faced were about losing their platform (and in-turn, their livelihood) due to changes of their platform's terms of service and being censored on their platform which the creators protected themselves by diversifying the platforms they were on and tracking what platforms required of self-censorship in order to comply as well as utilizing algospeak, which is a "deliberate misspelling, or substitution of words to avoid automated content moderation system" [9]. Sexual content creators also faced the risk of toxic content like harassment and hate, which they protected themselves by establishing clear rules about toxic content within their spaces and blocking any offenders of these rules [9]. Sexual content creators also were concerned about the risk of content leakage, or the distribution of their paywalled content as well as catfishes, or users that create fake profiles in order to trick others into doing something [9]. These were protected against by intentionally limiting their popularity to protect their personal information and their non-sexual identity, watermarking their paywalled content, doing routine searches for any leaked content elsewhere on the internet, and sharing information about any impersonators with the real individual when an impersonator is discovered and claiming their username as soon as possible when they are considering different platforms to expand to [9]. Soneji et al. also asked the Onlyfans creators about how they discovered the safety practices they utilize, and the creators mentioned following different digital safety learning sources like youtube channels and from following keywords about platform specific safety practices [9].

F. What sort of common behaviors do privacy concerned individuals stay away from?

Liu et al., in their paper, "Impact of Social Media Behavior on Privacy Information Security Based on Analytic Hierarchy

Process”, performed a online survey on 421 teenagers about their behaviors regarding social media and their concern for privacy [11]. Liu et al. scored different social media behaviors based on more privacy-preserving and less privacy-preserving and found that even though most users are generally aware of privacy-preserving behaviors like routine password changes, they usually fail to implement them sufficiently [11].

III. METHODOLOGY

To help prevent any incidents from occurring due to weak information security practices on social media, the authors of this paper aimed to put together a list of simple, minimally-technical behavioral guidelines, specialized for VTubers that one can follow to improve their privacy and security on social media in order to lower the chances of becoming a victim of cybercrime. The specialization of these guidelines for VTubers addresses the problem of behavioral security guidelines being too general to be effective [4] by offering specialization of behavioral infosec guidelines for this specific population of individuals, coupled with the fact that members of this population are frequent users of SNS for brand advertising and marketing purposes and have a high propensity for being victims of cybercrime, requiring proper infosec knowledge and behavior in order to protect their security and privacy [5]

A. Guideline Creation

- Guidelines will be created using publicly available data from Twitch, Youtube, X, and Reddit. An observational analysis will be done to explore previous cases regarding breaches of vtubers’ privacy and security as well as current practices used by vtubers to protect their individual privacy and information security.
 - The behavioral guidelines added to the list will be made to avoid any possible pairing of individuals or security vulnerabilities.
 - There will be no interaction with individuals during the creation of behavioral guidelines. This study is purely observational.
- This study will also consist of a literature review into the latest academic research relating to vtubers, vtuber culture, individual privacy and information security best practices on social media. Guidelines will be created from the academic literature and added to the list in order to provide the latest academic insight in a simple and to-the-point manner.
 - Any guidelines sourced from academic research will be cited appropriately.
- Once the guideline is created, it will be categorized into sections for easy reference so individuals can easily navigate the list and view infosec improving behaviors by what specifically the individual suspects they could improve.

B. Justification

“Most companies’ sns security guidelines are too general to be effective” [4]

High SNS usage + Risk Propensity correlated with increased likelihood of being involved in cybercrime [5]

IV. INITIAL RESULTS

Based on current events and best practices according to scholarly text, a list of guidelines has been compiled in order to assist content creators, and especially VTubers protect their privacy and safety both online and offline.

There have been past incidents of VTubers receiving Apple Airtags hidden in their mail and gifts, which sends location data through nearby iPhones in order to allow it (and any attached items) to be found. This Airtag allows the gifter to learn the potential whereabouts of a content creator’s home, and potentially lead to stalking. Incidents such as these, as well as can be very informative when it comes to the procedure a content creator should use when collecting and opening fan mail. First, if the content creator plans to allow fan mail, they should get a P.O. box in order to keep their address private. Second, they should open all of their mail and gifts where they received them, and search them thoroughly to ensure that there is nothing of potential harm to them included, such as a tracking device. Content creators can also use software to help detect tracking devices, such as AirGuard, which reports on nearby trackers, even if they’ve been modified to be stealthier. [6]

On that note, stalkers can be a threat even without the use of tracking technology. Indeed, many streamers fear or have experienced being doxxed, stalked, or having strangers show up to their home according to P. Samermit et al. [10] Because VTubers are typically live performers who exist through the internet, they often are sharing their screen with viewers, whether it be to play a game, demonstrate artistic skills, or just to serve as a visual aid to a conversation. While screen sharing during a livestream, it’s important to have vigilance about what might be shown. Because of this fact, actions that can be taken to prevent personal data from being exposed can include closing windows and tabs that could reveal names or addresses, and minimizing information shown on websites that could jeopardize privacy. According to Redditors that populate the r/Twitch community, PayPal in particular can be a big risk because it may show up due to incoming donations and it will often have one’s full name. [7] This risk can be reduced by giving the PayPal account a business name instead. Some other actions that can be taken to reduce the risk of stalkers include purging data broker information with online tools, keeping your location and surroundings secret, avoiding posting too many personal plans, and posting fake information.

Another risk that VTubers might encounter is impersonation. There are many reasons why one might want to impersonate a VTuber, but among them is the desire to ruin an individual’s reputation, or to abuse the trust that that content creator has in order to more successfully execute scams. On Youtube and other platforms, fake accounts may attempt to gather information through phishing. [8] To avoid impersonators, a content creator might blacklist words that are commonly associated with these impersonation scams from

appearing in their chat or comments. Making followers aware of impersonators may also be helpful, because it can help them catch on to these types of attacks and report the accounts behind the attacks.

V. DISCUSSION & CONCLUSION

The following bullets will be discussed in the final draft of this paper

- Discuss how parasocial relationships can commonly form among fans of VTubers, and the harm and discomfort it can cause. Include methods VTubers use to ward off strong parasocial relationships while remaining friendly or in character.
- Discuss talent agencies, and how they affect how a VTuber is able to maintain privacy and security.
- Conclude by restating why VTubers in particular need strong guidelines when it comes to their privacy and security (problem statement). List the guidelines, and briefly explain how they protect the content creator.
 - Discuss that even though VTubing is more technologically complex than normal streaming, VTubers are not necessarily infosec + security savvy.

REFERENCES

- [1] A. Turner, "Streaming as a Virtual Being: The Complex Relationship Between VTubers and Identity," Jun. 2022. Accessed: Sep. 14, 2024. [Online]. Available: <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1676326&dswid=-9638>
- [2] J. Byron, "New collaboration in a virtual world: studying Vtubers through identity, gender and fan engagement," figshare, Mar. 2023, doi: <https://doi.org/10.25949/22197799.v1>.
- [3] A. Tonami, M. Yoshida, and Y. Sano, "Online harassment in Japan: Dissecting the targeting of a female journalist," *F1000Research*, vol. 10, p. 1164, Feb. 2022, doi: <https://doi.org/10.12688/f1000research.74657.2>.
- [4] M. Silic and A. Back, "The Dark Side of Social Networking Sites: Understanding Phishing Risks," *SSRN Electronic Journal*, vol. 60, 2015, doi: <https://doi.org/10.2139/ssrn.2634887>.
- [5] G. Saridakis, V. Benson, J.-N. Ezingeard, and H. Tennakoon, "Individual information security, user behaviour and cyber victimisation: An empirical study of social networking users," *Technological Forecasting and Social Change*, vol. 102, pp. 320–330, Jan. 2016, doi: [10.1016/j.techfore.2015.08.012](https://doi.org/10.1016/j.techfore.2015.08.012).
- [6] T. Hopkins and N. Hutchins, "Apple's AirTag: A Security Vulnerability Discussion and Guide to Personal Safety," 2024.
- [7] "How to prevent being swatted?," Reddit.com, 2024. <https://www.reddit.com/r/Twitch/s/aJNxtgKJar> (accessed Nov. 11, 2024).
- [8] O. Goga, G. Venkatadri, and K. P. Gummadi, "The Doppelgänger Bot Attack," *Proceedings of the 2015 ACM Conference on Internet Measurement Conference - IMC '15*, 2015, doi: <https://doi.org/10.1145/2815675.2815699>.
- [9] A. Soneji, V. Hamilton, A. Doupé, A. McDonald, and E. M. Redmiles, "'I feel physically safe but not politically safe': Understanding the Digital Threats and Safety Practices of OnlyFans Creators," presented at the 33rd USENIX Security Symposium (USENIX Security 24), 2024, pp. 1–18. Accessed: Nov. 11, 2024. [Online]. Available: <https://www.usenix.org/conference/usenixsecurity24/presentation/soneji>
- [10] P. Samermit et al., "'Millions of people are watching you': Understanding the Digital-Safety Needs and Practices of Creators 'Millions of people are watching you': Understanding the Digital-Safety Needs and Practices of Creators," 2023. Accessed: Nov. 12, 2024. [Online]. Available: <https://www.usenix.org/conference/usenixsecurity23/presentation/samermit>
- [11] Y. Liu, W. K. Tse, P. Y. Kwok, and Y. H. Chiu, "Impact of Social Media Behavior on Privacy Information Security Based on Analytic Hierarchy Process," *Information*, vol. 13, no. 6, p. 280, May 2022, doi: [10.3390/info13060280](https://doi.org/10.3390/info13060280).