SecurityTok: Evaluating Pro-Security Advice on TikTok

Avi Gupta
Stanford University

Griffin Miller
Stanford University

Abstract

In this work, we study the quality of modern technological security education on social media. We focus specifically on TikTok, a dominant social media platform on which users can post short-form videos with entertaining content such as dance routines, product reviews, comedy skits, and much more. Using a methodology derived from recent and related literature [11], we collected 60 videos on TikTok across twelve categories of security advice to understand (a) how popular pro-security advice is on the platform and (b) the quality of this advice in terms of perceived actionability, perceived efficacy, and comprehensibility. We generally find that pro-security content, which has reached millions of viewers, is reasonably actionable and effective. However, we also discovered that some security advice disseminated on the platform can be dangerous. Based on these findings, we make recommendations to TikTok, and more generally, to social media technology companies.

1 Introduction

TikTok is a surprisingly flourishing platform for pro-security advice. Creators have shared content that describes, for example: how to create strong passwords, what to do if a computer gets hacked, and how to watch for suspicious phishing links. We call such videos "pro-security advice", as opposed to "anti-security advice" as coined in previous literature [13], because they appear to involve techniques that should improve a user's security, safety, and/or privacy. These videos are widely engaged with on the platform; the average video from our novel dataset of 53 pro-security TikToks has 967,605 views, 70,044 likes, 463 comments, and 14,189 shares. As an example, consider the following transcript of a one-minute and four-second TikTok (with 15,100 views) posted by @cybersecuritygirl. The video's speaker, Caitlin, says:

> Here are my top three tips to prevent you from getting hacked. Hi, I'm Caitlin. I'm a cybersecurity and data protection expert and I want to help you not get hacked. Let's go! The first: eliminate weak passwords. Instead, use a password generator like Keeper to create unique passwords for each account and help prevent compromised credentials. Two: use multi-factor authentication or MFA. MFA is when you need two or more special codes or things like a password and a fingerprint to make sure it's really you logging into your account. And many cloud solutions offer MFA – which can prevent 99.9% of password-related cyber attacks on your account according to Microsoft. And third: delete inactive accounts. Since a cybercriminal could use an inactive user account, keeping an account alive but inactive is a crucial security risk. (TT17)

Videos such as this one have the potential to reach and influence any number of TikTok's nearly 700 million users worldwide [3]. U.S. adults spend on average 46 minutes per day on TikTok according to a 2022 survey [6], and an estimated 37.3% of users on TikTok are between the young and impressionable ages of 18 and 24 [4]. It is therefore imperative that pro-security content on the platform is comprehensible, practical, accurate, and indeed risk-reducing.

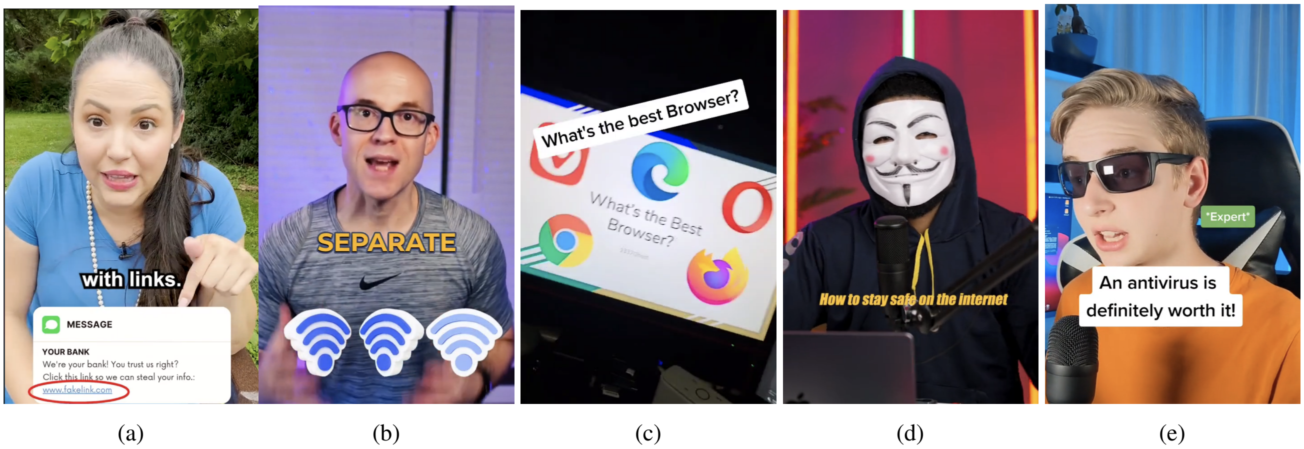
By studying the most highly viewed security advice on the platform at this current moment, we develop an understanding of the patterns in security imperatives that are ingested by viewers around the world and consider whether this advice truly aligns with security best practices.

2 Related Work

In this section, we outline some of the recent literature in the sphere of online advice.

2.1 TikTok and Social Media

In 2022, researchers at the University of Washington analyzed TikTok for videos related to anti-privacy, honing in on surveillance and abuse in parent-child relationships and intimate part-


with links.
MESSAGE
YOUR BANK
We're your bank! You trust us right?
Click this link so we can steal your info.:
www.fakelink.com
SEPARATE
What's the best Browser?
What's the Best Browser?
How to stay safe on the internet
Expert
An antivirus is definitely worth it!

(a) (b) (c) (d) (e)

Figure 1: Examples of Anti-Privacy and Anti-Security Advice on TikTok. a) Advice to protect against financial scams (TT32). b) Top 3 tips for a safe home network (TT25). c) Comparison of popular browsers (TT39). d) How hackers can gain access to devices through phishing (TT7). e) Comparison of popular anti-virus software (TT10).

ner relations [13]. The authors determined that anti-security advice in parent-child relationships was not framed as deviating from social norms and, rather, as protective and helpful. They discussed TikTok culture feeding into the creation of these videos, with an emphasis on strong emotional appeal, multi-modal content of music and speaking, and the interest in gaining views. The paper paves the way for future research into the rich abundance of videos hosted on the platform and encourages both an exploration into other subcommunities and mitigating the spread of misinformation through the algorithm. In fact, the authors provide the conceptual inspiration for this paper by writing that "future work might investigate pro-security advice on TikTok."

Multiple studies have been conducted regarding health and social media. Zenone et al. investigated health advice on TikTok connected to COVID-19 and sexual education and found that the quality of information and qualifications of the creators are largely unknown [14]. Specifically in a case study of acne medical information, they discovered “serious to potentially important shortcomings.” Martin Engebretsen discusses the balance of health/medical influencers creating engaging content to maintain their online presence (e.g. promoting resharing, commenting, and liking) and providing substantive advice for younger audiences [9]. Additionally, there have been similar studies connected to financial planning on social media. Bryan Teoh Phern Chern observed a transition away from human personal financial advisors to online advice and ultimately argues for higher regulation of the industry [12].

2.2 Security Advice

In their USENIX 2020 Distinguished Paper Award-winning paper, “A Comprehensive Quality Evaluation of Security and Privacy Advice on the Web,” Redmiles et al. measure the quality of security advice in online articles [11]. They make three primary contributions: creating a taxonomy of advice imperatives, developing measurement approaches for advice quality metrics, and subsequently using their framework to evaluate their dataset of end-user-focused security and privacy advice on online articles. The authors measure advice quality by scoring the “comprehensibility, perceived actionability, and perceived efficacy” of security advice on various forms of Likert scales. The authors ultimately find that the majority of advice is perceived to be at least somewhat comprehensible and actionable, but that users struggle to prioritize amongst the sheer volume of advice. We take heavy methodological inspiration from this paper by using a similar evaluation framework and set of security categories. [11].

Lorenzo et al. set out to better understand the cause of security advice overproduction in a semi-structured interview study with 21 advice writers [10]. They learn that authors attempt to cover a large amount of content but with few attempts to deprioritize or curate less essential content and only review or update content after major security events [10]. TikTok users, especially those who are keen on viewing pro-security content, may experience a similarly overwhelming amount of advice on their feed. However, TikTok videos are on average about 32 seconds in length, limiting the amount and depth of content that can be communicated in any one video. TikTok viewers seem to have finite attention spans and are quick to scroll to the next video. This dynamic may push creators to curate their tips and present them in a digestible fashion.

3 Methodology

We describe our methodology in three phases: creating our dataset of TikToks (3.1), tracking the videos’ relevant metrics (3.2), and scoring them (3.3). We additionally check our video evaluations against an inter-rater consistency metric (3.4).

3.1 TikTok Dataset

We constructed our dataset using videos that we searched for and selected directly from TikTok. To guide our search, we first selected twelve prevalent security topics to consider in our study based on previous literature. The categories are: passwords, account security, browsers, general security, antivirus, software, network security, device security, privacy, data storage, incident response, and finance security. [11]. Note that we interpret "passwords" as a subset of "account security" in that account security videos may mention techniques for password security, but must also mention at least one other technique for account security.

To create search queries for each security category, we could not always enter the name of the category itself. For example, searching for "financial security" on TikTok yields results geared towards videos providing advice for financial *stability* in the economic sense, not the technological sense. Instead, to generate search terms that would effectively search TikTok's database, we employed combinations of the following strategies:

1. Append "advice" or "tips" or "security" to the category name, such as "network security tips" or "password security."

2. Search for related hashtags (e.g. #phishing and #antivirus).

3. Add or replace words to narrow down the category into a more specific one (e.g. "IOT Security" for device security; "multi-factor authentication" for account security.

4. Create a situation/narrative, e.g. "What to do if I got hacked?" for incident response.

For each of the twelve topics, we created five search terms using the strategies outlined above. We then searched TikTok using these terms and selected the top-ranked result. For context, TikTok's search tab shows the "Top Liked" video first, and TikTok states that the ordering of videos is determined by "the most relevant results" [5] (note that TikTok does not provide any further insight as to how the algorithm explicitly determines relevancy). They also note that "the hashtag page displays the videos that started the trend first, and then other popular videos relevant to the trending hashtag" [5]. Users also have the ability to sort the results by date, but we did not utilize this feature.

By selecting the top result for each of our sixty queries, we collected an initial set of sixty TikToks. However, not all of our search terms produced videos related to the security topic at hand and provided some form of advice. In seven cases, we discarded the top-ranked TikTok for the search term based on our judgment of irrelevance, leaving us with a final dataset of 53 videos. Although our dataset's size is relatively modest, it is on the order of the 98 videos collected by Wei et al. [13].

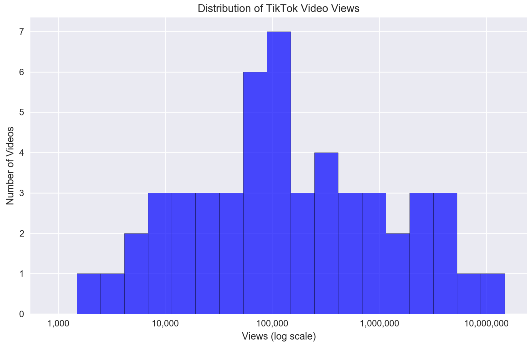


Figure 2: Distribution of Views

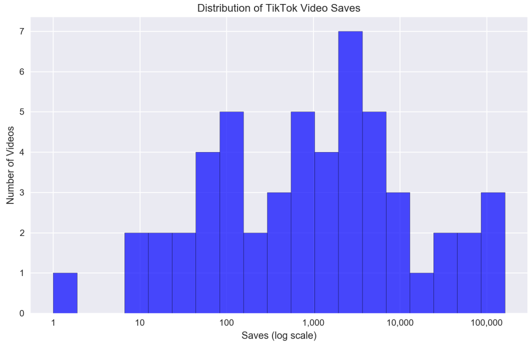


Figure 3: Distribution of Saves

The collection of 53 videos we use in the paper may not be consistent with a future study that uses the same video collection methodology because the top videos shown for each search term will change over time. Moving forward, we abbreviate the xth TikTok in our dataset to TTx, following the format of Wei et al. [13].

3.2 Video Statistics

TikTok surfaces several metrics on each of its videos, including the number of views, likes, comments, and saves. Each of these metrics provides insight into the user's reaction to the video: likes may suggest agreement, comments may indicate engagement, and saves may be a proxy for usefulness. Figure 2 shows the distribution of views among the videos we used in our dataset, ranging from 800 to 4 million views, with the average at around 967,000 views. As an example for another metric, Figure 3, similarly, shows the distribution of saves among the videos, ranging from 1 to 162,000 saves and an average of 14,000 saves.

3.3 Evaluation Metrics

To evaluate the quality of the security advice given in the selected TikToks, we devised a scoring procedure similar to the one used in Redmiles et al. [11]. Specifically, we considered three categories of metrics: perceived actionability, perceived efficacy, and comprehensibility. Perceived actionability is further described by confidence, time consumption, and disruption; these metrics consider the practicality of the advice. Perceived efficacy is described by risk increase, risk reduction, and prioritization; these metrics consider the impact of the advice.

- Perceived Actionability
  - *Confidence*: How confident are we that an average user would be able to implement this advice? We rank this on a scale of (1-4), with 1 being not at all confident and 4 being very confident.
  - *Time consumption*: How long do we believe it would take for an average user to implement this advice. We rank this on a scale of (1-4), with 1 being very little time and 4 being a lot of time.
  - *Disruptive*: How disruptive do we believe it would be for an average user to implement this advice? We rank this on a scale of (1-4), with 1 being minimally disruptive and 4 being very disruptive. Note: these scores were ultimately discarded, as explained in Section 3.4.
- Perceived Efficacy
  - *Risk Impact*: How much the risk of the user would change if the advice were followed? We rank the risk reduction from 0 (no change) to 50 percent (high-risk reduction), and similarly, the risk increase from 0 (no change) to 50 percent (high-risk increase). We do not assume that risk can either be exclusively increased or decreased, so it need not be the case that one of the two risk impact values is zero.
  - *Priority*: How highly would we recommend prioritizing this advice? We rank this on a scale of (1, 3, 5, 10, >10, and Would Not Recommend), with 1 being the strongest piece of advice for that particular category, 3 in the Top 3, etc.
- *Comprehensibility*: How easily do we believe this video will understood by an average user? We rank this on a scale of (-2, 2), with -2 being hard to comprehend and 2 being very easy to understand.

3.4 Scorer Reliability

To ensure that the two authors of this paper were aligned in their scores, we considered the Kendall Rank Correlation Coefficient, also known as Kendall's Tau, as our inter-scorer consistency metric. Kendall's Tau is a nonparametric measure of association based on the number of ordinal concordances and discordances in paired observations. A concordance occurs when for any two videos, both scores change in the same direction (either both increase or both decrease). A discordance occurs when the scores move in different directions. Specifically, $\tau = C - D/C + D$, which is the normalized difference of the concurrences in discordances. Intuitively, the the coefficient ranges from -1 to 1 and will be higher when observations have a similar rank, and lower when they have a dissimilar rank.

Category	Kendall's Tau Correlation	P-Value
Confidence	0.59	0.004
Time Consumption	0.52	0.008
Disruption	-0.02	0.09
Risk Reduction	0.5	0.003
Risk Increase	0.24	0.284
Priority	0.7	0.001
Comprehensibility	0.51	0.007

Table 1: Kendall's Tau Correlation between Two Scorers

The consistency scores were calculated using an online tool, DataTab [8], which provides the functionality to input evaluation scores for each metric and view both the Kendall's Tau correlation as well as the statistical significance. The null hypothesis for the significance test was: there is not a positive correlation between the rankings of the two scorers. Based on our results, we can reject this null hypothesis at a p < 0.05 level across all metrics except two: disruption and risk increase. We see moderately strong positive linear correlations in all metrics except the same two metrics. We moved forward with the data in the risk increase because the correlation is still positive, but discarded the disruption evaluation metric because the correlation was weakly negative. From here on out, we only consider our confidence and time consumption evaluations in the actionability category.

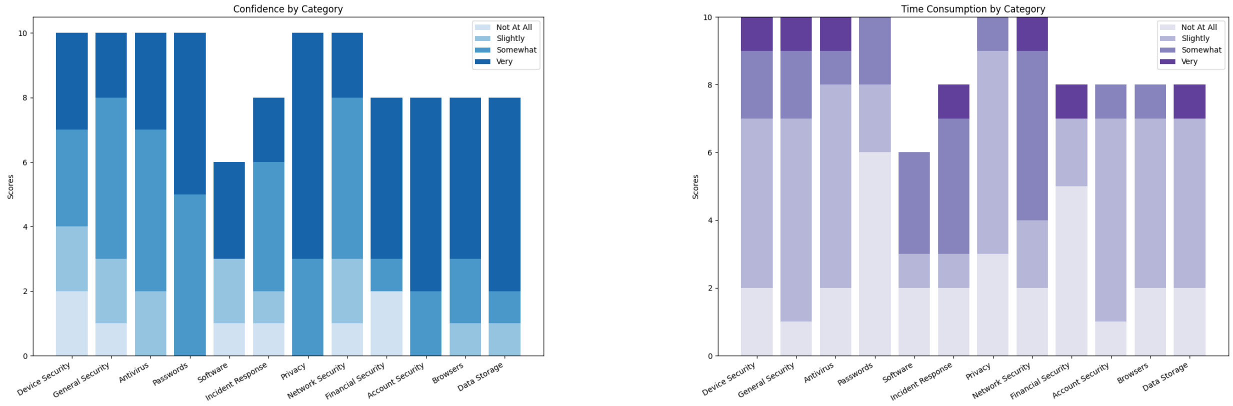
4 Results and Findings

4.1 Perceived Actionability

In this section, we summarize our evaluations of the perceived actionability of the 53 pieces of advice in our corpus. As described previously, we break down perceived actionability into two categories: confidence and time consumption.

4.1.1 Overall Actionability

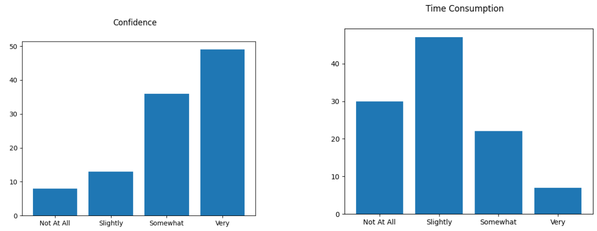
We find that the majority of advice found on TikTok is reasonably actionable. Using a concatenation of the scores from both raters, the median rating for confidence was 3 out of 4



(a) Confidence Score Counts by Category

(b) Time Consumption Score Counts by Category

Figure 4: Confidence and Time Consumption Score Counts by Advice Category



(a) Confidence

(b) Time Consumption

Figure 5: Distributions of Confidence and Time Consumption

("somewhat" confident). The median time consumption score was 2 out of 4 ("slightly" time-consuming). The distribution of confidence and time consumption scores given by both raters are shown in 5a and 5b respectively. The raters were "very" or "somewhat" confident in the practicality of eighty percent of videos (see Figure 5a), and scored 73% of videos in the "not at all" or "slightly" time-consuming buckets (Figure 5b).

4.1.2 Actionability By Topic

Figure 4a is a stacked bar chart showing the relative confidence scores given in each security category. Privacy reigns when it comes to the confidence scores, with seven videos receiving scores of "very" confident that an average viewer could implement the provided advice. Many of these videos gave step-by-step instructions for disabling location tracking and data sharing features on commonly used interfaces such as the iPhone's Settings application or the Venmo application, and the scorers most likely agreed that these processes were achievable by the average user. Passwords also have notably high scores for actionability confidence, with 50% of scores at the "somewhat" confident level and the other 50% in the "very" confident level (and therefore none in the lower two tiers). The categories with more questionable actionability include antivirus and incident response. We saw that some of these videos suffered from being inaccurate or outdated. For instance, one TikTok that explains how to recover a hacked Roblox account (in the incident response category) describes a method of contacting customer support, but several comments on the TikTok indicate that the support button did not appear for them. The raters scored this particular video an average of 2.5 out of 4 in confidence.

Figure 4b visualizes the relative scores of time consumption by each category of security advice. With one exception, all categories of advice received the most scores of either "not at all" or "slightly" time-consuming. The categories with the most scores in the "not at all" time-consuming bucket were passwords (6) and financial security (5). As a concrete example, advice given in the financial security category included being aware of financial scams from people pretending to be a bank, which requires awareness but is not very time-consuming. Privacy comes in at a close second, with 3 scores in the "not all all" level but 90% of its scores in the lower two buckets of time consumption. The exception was network security, which received score counts of [2, 2, 5, 1], meaning that the most frequently scored level was the "somewhat" time-consuming bucket. This may be explained by the complexities associated with networking device interfaces, such as home routers, which are traditionally not the most user-friendly. For example, several videos suggested setting up a network firewall, but most non-security experts may have difficulty doing so for the first time, or at the very least, will need to research the available options and execute the setup procedure.

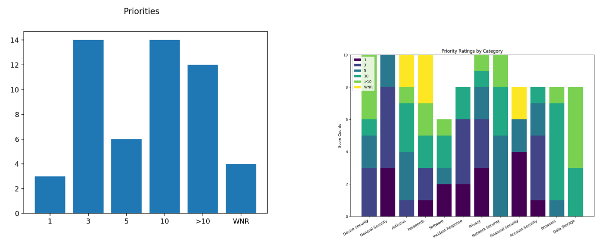
4.1.3 Extrapolations from Actionability Scores

Considering both our confidence and time consumption scores, we conclude that **privacy may be the most actionable category of advice on TikTok**, with top-skewed confidence scores ([0, 0, 3, 7]) and bottom-skewed time consumption scores ([3, 6, 1, 0]). The higher time consumption scores for the network security category are balanced by the higher confidence scores, leaving its actionability somewhere in the middle. There does not seem to be a clear "loser" of actionability advice, but the device security category did have a very even spread of scores in all levels of metrics ([2, 2, 3, 3] for confidence and [2, 5, 2, 1] for time consumption). This category may be impacted by the broad definition of "device"; in our dataset, we included several kinds of mobile devices and IOT devices in this category. Additionally, the incident response category stands out for both its relatively higher (4, compared to mostly 1s or 2s) number of scores in the "somewhat" time-consuming level and its relatively low (2, compared to mostly 3s, 5s, and 6s) number of scores in the "very" confident level.

4.2 Perceived Efficacy

In this section, we summarize our evaluations of the perceived efficacy of the 53 pieces of advice in our corpus. Recall that we break down perceived efficacy into priority and risk, as described in Section 3.3.

4.2.1 Priority



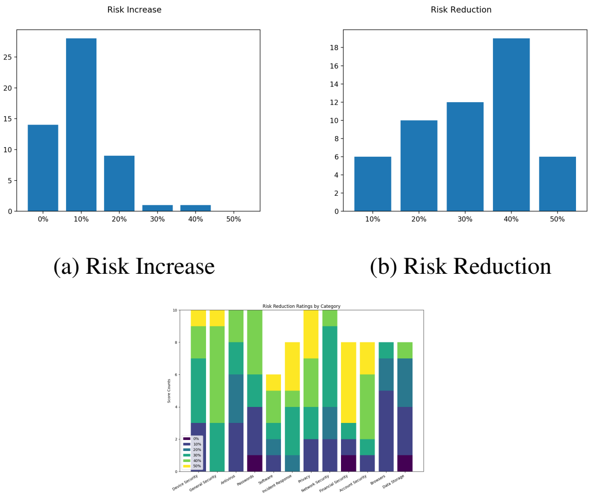
(a) Distribution of Priorities

(b) Priority Counts by Category

Figure 6: Priority Score Distribution and Counts

Overall, the majority of TikToks were scored as in the Top 3 or Top 10 priority levels, as seen in Figure 6a. 69.8% of all videos fell within advice prioritized within the Top 10 level (including the Top 1, Top 3, and Top 5 buckets), which is a positive sign of the quality of security advice on the platform. As an example of a high-priority video, both authors ranked Priority 1 for a TikTok urging users to turn off their precise location on the dating app Grindr (TT24). Figure 6b suggests that **general security was the category with the highest proportion of high-priority security advice**, with 3 scores of Top 1, 5 scores of Top 3, and 2 scores of Top 5. Some of these security tips included taking caution before clicking on hidden URLs on Discord (TT6), how to watch for and avoid phishing attacks (TT7), and a reminder to never include API keys in public code repositories (TT46). Privacy and account security also emerged as strong categories for high-priority advice. The category with the lowest priority advice was data storage and browsers, with only one score in the Top 5 across both categories.

4.2.2 Risk



(a) Risk Increase

(b) Risk Reduction

(c) Risk Reduction by Category

Figure 7: Scores for Risk Change

We found that the majority of videos increase security risks by a maximum of 10% as seen through Figure 7a. An example of a video with a 10% increased risk is shown in TT9, where the creator encourages Windows users to do antivirus scans. While the advice is sound, it is not a comprehensive solution. There still remains a risk that users might assume they are virus-free if they use antivirus software that fails to detect certain malware. We also found that the videos most commonly decreased security risks by up to 40%, as shown in Figure 7b. We interpret these findings as a positive indicator in terms of the overall quality of advice being disseminated, as it implies that most content does not substantially endanger users' security and that there exists valuable advice that would enhance user security if implemented. Importantly, however, **few videos are associated with a very high potential increase in security risk**, indicating the presence of videos containing particularly bad advice, misleading information, or harmful practices. For example, TT45 posted a link to a suspicious website claiming that it contained the best malware detection software and encouraged the viewers to download the executable via their terminal. Though the integrity of this link is unclear, without loss of generality the payload of the download could potentially be harmful.

We briefly turn to Figure 7c to consider risk reduction at a category level. We find that (similarly to our priority score results), browsers and data storage do not boast high-risk reduction whereas general security does. This relationship between priority scores and risk reduction scores is expected, given that higher risk reduction should be at least a key factor in the priority of a piece of security advice. On the risk increase side, category-level analysis did not yield interesting results as the relatively few videos with high risk did not trend toward any particular category.

4.3 Comprehensibility

We find that the median comprehensibility value from our dataset is 1.5 (derived from discrete scores ranging from -2 to 2). The nature of the TikToks we are searching for are inherently advice and explanation-driven, so the high median comprehensibility is not surprising. Although we did not explicitly score any sub metrics more detailed than "comprehensibility," videos that were particularly comprehensible tended to use slower speech paces, presented the security advice in a structured form (rather than speaking in a stream-of-consciousness manner), and utilized a relevant combination of audio and visuals (static or dynamic). Expanding on the last point, we noticed several videos that presented advice by only showing text on the screen with no voiceover or using an audio file with a completely irrelevant visual on the screen. In these cases, the lack of one of the two communication media (auditory or visual) detracted from the accessibility and thus comprehensibility of the videos. The videos with lower comprehensibility scores did not trend towards any one security category but rather fell flat on the way that the information was presented. For example, in TT15, a password security advice video, example passwords, and their respective approximate hacking times were displayed in a small text square on the screen with no speech audio and a background video of a person typing on a computer. Without audibly relevant information and legible text, the median comprehensibility score of the video was -1.5. Comprehensibility can be more important than it initially seems; misread, misheard, or misinterpreted security advice may have detrimental security implications for viewers.

5 Relationships Between Metrics and Advice

As an additional step in our analysis, we investigate the potential relationship between the videos and their popularity. We first analyzed the relationship between views and priority levels as seen in Figure 8a. A Spearman's rank correlation coefficient of 0.301 indicated a positive but relatively weak relationship between TikTok views and the encoded priority rankings. This similarly reflected saves, as shown in Figure 8b, with a coefficient of 0.302. This trend continued with a comparably low correlation between likes and comments with priorities, yielding coefficients of 0.313 and 0.216, respectively. The findings might suggest that user engagement is not necessarily a function of the quality or priority of the security advice. Instead, users may engage with content for various reasons, which could include entertainment value, as further discussed in Section 6.2.1.

We then examined the relationship between the video statistics and the security risk increase evaluation metric. As shown in Figure 9a and Figure 9b, there appears to be a cluster of videos with moderate views and saves that have a varied range of risk increase scores. This could suggest that content with moderate popularity covers a non-trivial range of security advice quality. Videos with the lowest risk increase scores are spread across the entire range of views and saves, indicating that safe advice can either be very popular or relatively unnoticed.

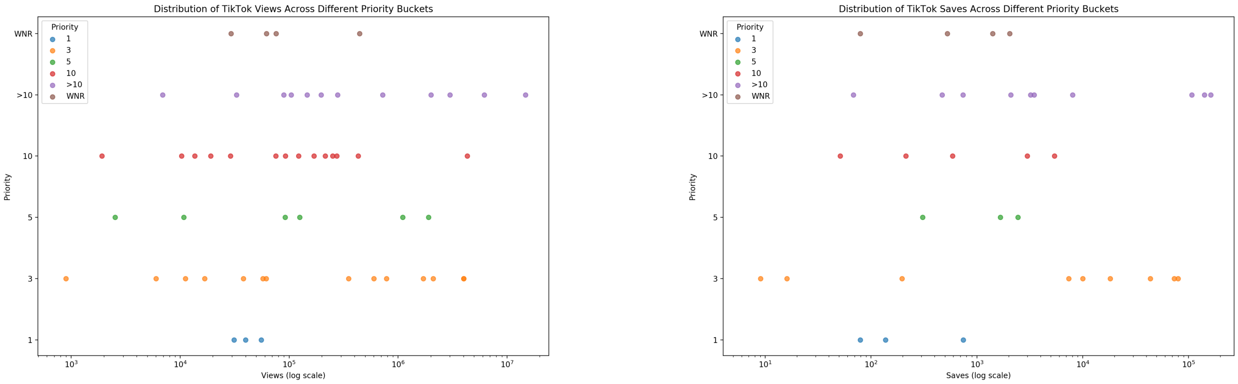
Finally, we explored the connection between TikTok's video metrics and the video content's actionability attributes. The confidence metric produced weakly inverse correlations when examined against views, likes, and saves, yielding Spearman correlation coefficients of -0.206, -0.225, and -0.233, respectively. This finding is certainly unexpected, as we would expect more actionable videos, for example, to receive more saves. The result we see may speak to some bias in our confidence scores or could suggest that saves are not the best proxy for confidence of practicality. Similarly, we found weak correlations between time consumption and the previously mentioned metrics, yielding coefficients of 0.179, 0.109, and 0.234. These results show a weakly positive correlation between the number of saves, views, or likes that a video receives and the time cost of the advice it describes.

6 Discussion

6.1 Limitations

Our discussion of category-level results may be impacted by our decision to remove seven irrelevant videos in our dataset. As a result of this decision, certain categories did not receive five videos. These categories may therefore be at a disadvantage when we compare categories against each other. If we were to do this study again, we would search for replacement TikToks in their respective categories to allow for even inter-category comparisons.

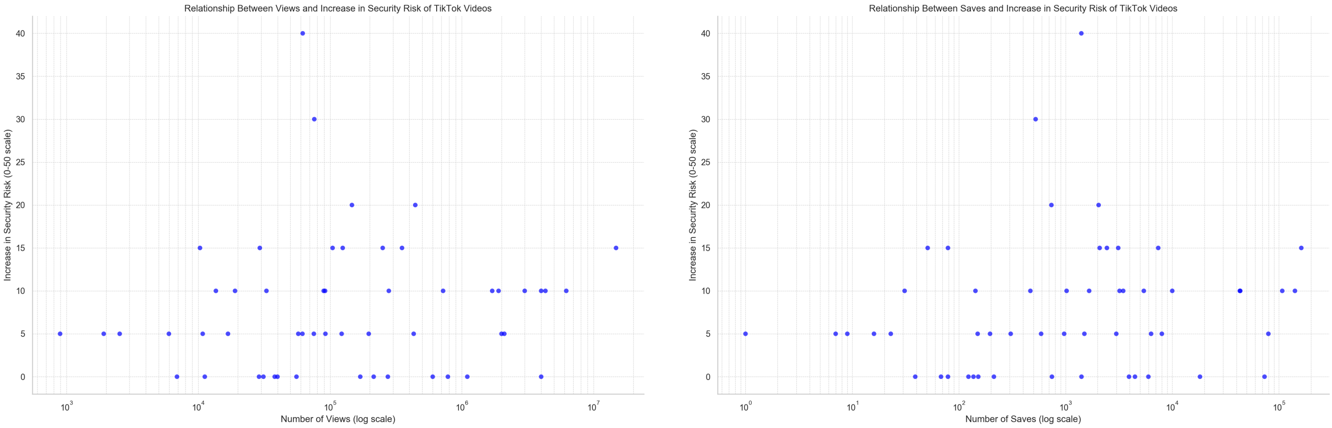
Our category-level conclusions may have been impacted by our personal biases towards the security category itself, and not necessarily the specific advice given each video. For example, in the case of the priority metric, we may have unintentionally scored a video partly based on our impression of the priority of its security category, and not only the priority of the specific piece of advice given in the video. Additionally, our ratings may be biased by the background of the two scorers, who have received highly similar undergraduate and graduate education at the same university.



(a) Views against Priority

(b) Saves against Priority

Figure 8: Comparative Analysis of Priority Correlations



(a) Views against Security Risk

(b) Saves against Security Risk

Figure 9: Comparative Analysis of Security Risk

Our search terms, which were generated in a rather arbitrary manner, are also certain to have significantly impacted our downstream evaluations of the quality of advice on TikTok. For instance, using "phishing" as one of our five search queries for the general security category instead of, say, "avoid partner abuse," may have produced advice with lower time consumption, which in turn could have led to a hypothetical conclusion that general security was the least time-consuming subject. Furthermore, the search terms we used in any given category, and the relatively low number of videos we were able to manually process, are unlikely to represent the extent of security videos in that category. We would ideally have used many more search terms and many more videos in our analysis.

6.2 Qualitative Observations

In this section, we describe some of our qualitative observations from the pro-security videos we found on TikTok.

6.2.1 Incentive Misalignment

A key qualitative observation from our study is the potential misalignment of information quality and financial incentives for video creators on TikTok.

Creators on TikTok are financially incentivized to create videos that receive more views. Views are determined by TikTok's virality algorithm, and videos that go viral tend to have some degree of entertainment. TT32 features a verified creator, which TikTok grants a blue check-marked badge to accounts that are "active, authentic, notable, and unique" [7]. The creator edited sound effects, used props, and used graphics and animations to display text, creating a highly entertaining and captivating video with over 55,000 views and 8,000 likes. In TT17, another verified creator presents research about an artificial intelligence model that "hears" passwords through keyboard click sounds. This creator uses the popular filter green screen to display herself over the paper, GIFs and closed captions to create an entertaining video, employing methods from other creators for any subject. This video gathered over 32,000 views and 3,000 likes.

There are two sides to this coin: on one hand, the security advice on TikTok is often entertaining, making it digestible for the average person. On the other hand, viewers may be less critical of this advice if it is presented in an entertaining manner or by a "verified" account, and creators may feel pressured to shorten their videos or even compromise on the accuracy of their statements in an effort to maintain the video's entertainment value and hold their viewers' attention.

Sponsorships represent an additional layer of financial incentives for creators. We observed that several of the videos in our dataset seemed to include sponsored content. The @cybersecuritygirl video described in this paper's introduction section ends with:

> To help you address tip #1, use Keeper security. It's a solution that you can use to protect your passwords from cyber criminals and prevent password-related data breaches. So take advantage of 40% off either a personal or family plan for the first year using Cybergirl40 from now until June 23rd. (TT17)

Although Caitlin does not say it explicitly, this video was likely sponsored by Keeper. Presumably, sponsors will tend to choose creators – like Caitlin – who are verified and have a trusted base of followers. However, these creators may not as rigorously vet their corporate sponsors, and run the risk of unintentionally recommending a malicious product to their user base.

6.2.2 The Comment Section

The comment section provides a forum for creators and viewers to interact with one another once a TikTok has been posted. We noticed that creators will sometimes use the comment section to make clarifications on their video content. In a video discussing the benefits of ledgers for securing cryptocurrency (TT4), the creator comments "something I should have added is that you should only purchase a ledger from their official outlets. Third party hard wallets are not secure." This is the second most-liked comment on this discussion of 130 comments with 35 likes. Although the creator makes this point eventually, it might have been too late for someone to fall victim to a ledger scam.

Even more commonly, the comment section allows viewers to inform others if the information in the video may be inaccurate or out of date. In a TikTok from our dataset that describes how to launch a diagnostic menu on an Android phone (TT2), many comments complain that it is difficult to exit the menu without restarting the phone. We noticed this when scoring the video, and lowered our actionability scores accordingly. Another TikTok that explained how to contact Instagram customer service to repair a hacked account (TT20) was repeatedly noted to be out of date in the comment section, as the "customer service" button was no longer available as it was described.

Although the comment section can be a wonderful tool to flag videos that have lower qualities of advice or misinformation, commenting also increases that video's engagement scores, which may boost the video's popularity in TikTok's virality algorithm. TikTok does have a "report" button, but it is unclear how responsive TikTok is to reports.

One of the most shocking videos we found was a video with the text "Top 4 Passwords to Use" followed by four password strings (TT14). This video was most likely created as a joke, but there is no explicit indication of this anywhere. Therefore, many of its 75,900 viewers almost certainly have adopted the exact passwords suggested by the video, leaving them vulnerable to hackers who see this TikTok. One of the most-liked (219 likes) comments on the video affirmingly says, "Wow they work great, thanks." This comment was likely made sarcastically, but not everyone may not read it as such. Unfortunately, such comments on harmful videos can potentially boost the video's popularity and reputation.

6.3 Content Moderation

TikTok's approach to content moderation combines automated moderation technology with manual human moderation to identify content that violates their Community Guidelines [2]. Specifically, their guidelines protect against videos that "contain policies on removing harmful misinformation that could mislead [the] community about civic processes, public health, or safety" [1]. Their moderation process uses 15 fact-checking global organizations, supporting more than 40 languages. They additionally flag indeterminate videos with banners to alert viewers.

As we discovered, poor security advice is present across TikTok, and we encourage TikTok to continue developing their moderation algorithm and suggest they ensure their fact-checking organizations have collaborated with security professionals to prevent harmful security advice from circulating. We want to stress that it is important that these moderation policies affect videos of all view counts, not just popular ones or videos posted by verified creators.

7 Future Work

We believe future work should first and foremost study this subject on a larger scale, with more videos and different categories. Additionally, raters from diverse backgrounds such as education and age could produce more reliable results.

To understand TikTok's algorithmic behavior, future work could examine the time taken to remove or flag videos with poor advice. Researchers could also analyze the TikTok virality algorithm for spreading content related to pro-security advice. It would additionally be interesting to see how security advice compares between different social media platforms such as Instagram, Facebook, YouTube, and X in terms of the quality and impact of security advice.

For the scope of this project, we we unable to interact with people. Future researchers could conduct interviews with TikTok users to study their decision-making processes in regards to trusting and implementing security advice from social media as well as interacting with the videos through likes, comments, and saves. The quality of a video may be more accurately measured by asking a panel of security experts if they would adopt the advice provided in the video.

Finally, it would be of utmost importance to investigate the ethical implications of security advice on social media, and further research may propose methods to ensure the spread of security-related advice on social media is held to modern ethical codes.

8 Conclusion

The popularity of TikTok among other social media platforms has been rapidly increasing over the past decade. As a fast source of information, average users may turn to social media platforms for security advice, and social media platforms must anticipate this and ensure the quality of advice disseminated is high. In this study, we find that privacy advice videos are particularly actionable, general security advice videos are often high-priority and not very time-consuming, and most pro-security advice videos are fairly comprehensible. However, four of the 53 scored videos in our corpus (7.5 percent) had priority scores of "Would Not Recommend," as their security imperatives were either futile or actively harmful to viewers' security, despite seeming like legitimate pro-security advice at first glance. Therefore, we strongly suggest that TikTok viewers act on pro-security advice with cautious optimism. They should avoid blindly trusting creators, even if they are "verified." TikTok should ideally issue frequent reminders to users to be aware of misinformation on the application. TikTok should also improve their content moderation of pro-security videos on their platform in consultation with security experts who can advise on how to best identify security misinformation.

References

[1] Combating misinformation. https://www.tiktok.com/transparency/en-us/combating-misinformation/. Accessed: 2023-12-03.

[2] Our approach to content moderation. https://www.tiktok.com/transparency/en-us/content-moderation/. Accessed: 2023-12-03.

[3] Tiktok - statistics. https://www.statista.com/topics/6077/tiktok/. Accessed: 2023-12-04.

[4] Tiktok age demographics. https://www.oberlo.com/statistics/tiktok-age-demographics. Accessed: 2023-10-15.

[5] Tiktok discover and search page. https://support.tiktok.com/en/using-tiktok/exploring-videos/discover-and-search. Accessed: 2023-11-17.

[6] U.s. time per day on netflix, tiktok, youtube 2024. https://www.statista.com/statistics/1359403/us-time-spent-per-day-netflix-tiktok-youtube/. Accessed: 2023-10-15.

[7] Verified accounts on tiktok. https://support.tiktok.com/en/using-tiktok/growing-your-audience/how-to-tell-if-an-account-is-verified-on-tiktok. Accessed: 2023-12-02.

[8] DataTab. Kendall's tau calculator, 2023. Accessed on: 12/1/2023.

[9] Martin Engebretsen. Communicating health advice on social media: A multimodal case study. *MedieKultur: Journal of media and communication research*, 39(74):164–184, May 2023.

[10] Lorenzo Neil, Harshini Sri Ramulu, Yasemin Acar, and Bradley Reaves. Who comes up with this stuff? interviewing authors to understand how they produce security advice. In *Proceedings of the Nineteenth USENIX Conference on Usable Privacy and Security*, SOUPS '23, USA, 2023. USENIX Association.

[11] 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.

[12] Bryan Teoh. Evaluating the evolution of the personal financial planning industry: Mutualism, commensalism, or parasitism. *GATR Journal of Finance and Banking Review*, 6:72–81, 09 2021.

[13] Miranda Wei, Eric Zeng, Tadayoshi Kohno, and Franziska Roesner. Anti-Privacy and Anti-Security advice on TikTok: Case studies of Technology-Enabled surveillance and control in intimate partner and Parent-Child relationships. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*, pages 447–462, Boston, MA, August 2022. USENIX Association.

[14] Marco Zenone, Nikki Ow, and Skye Barbic. Tiktok and public health: a proposed research agenda. *BMJ Global Health*, 6:e007648, 11 2021.