Research Article

From Posts to Protection: Understanding User-Generated Safety Content on Reddit

Mashael Yousef Almoqbel*

Jubail Industrial College, Business Administration department, Assistant professor of Management Information Systems-Saudi Arabia
* **Corresponding Author Email:** moqbelm@rcjy.edu.sa- **ORCID:** 0000-0002-5247-567X

Abstract:

This study investigates discussions surrounding safety on Reddit to understand the topics and user comments related to safety concerns. By employing content analysis, the research explores the themes that emerge in posts and responses within safety-focused subreddits. The findings are interpreted through the lenses of Social Support Theory and Uses and Gratifications Theory, providing insight into how individuals seek information, express concerns, and interact with others about safety-related issues. The analysis reveals key themes such as personal safety, community vigilance, and the sharing of safety resources. Based on the findings, several design implications are suggested to improve the user experience and effectiveness of safety-related content on social platforms. These implications aim to enhance community engagement, foster supportive interactions, and better address user needs in digital spaces.

1. Introduction

Social media growth is increasing exponentially and reached 239 million US users in 2024 [1]. Although not in the top five most used social media platform, Reddit has roughly 365.4 million weekly active users worldwide with an annual revenue of around $872.5 million [2]. Reddit is a popular social media platform where users can post content, share links, and participate in discussions. It's organized into "subreddits," which are individual communities dedicated to specific topics, interests, or themes, such as news, technology, entertainment, or hobbies. Each subreddit has its own set of rules and culture.

Safety on social media is a well-studied topic, given that generations are becoming more and more comfortable sharing their thoughts with strangers online. Many research studies have focused on improving safety on social media by detecting unsafe behaviors such as harassment, cyberbullying, hate speech, and misinformation [3,4,5,6]. These efforts often use natural language processing (NLP) and machine learning techniques to automatically identify harmful content, such as aggressive language, threats, and offensive remarks [7]. Other research studied what attributes affect safety perceptions of social media users. For example, in a diary study, participants reported that communication with their real and virtual communities enhanced their safety perceptions [8]. In the latter study, some participants reported that certain features on social media apps that allow them to report incidents or unfavourable behaviours increase their feeling of safety. More recent work looked at how to mitigate online safety related issues such as bullying, harassment, and toxic metaphors [9]. Other work looked at how policies affect online content and how safe it is as well as its effect on self-censorship where authors would amend their comments and reactions to posts [10]. A lot of research studies moderation on social media to enhance safety for users [11,12,13].

In the context of sharing safety related content on social media, researchers studied reasons of social media users to share public safety incidents on social media, where participants who showed more altruistic traits, were more likely to share such incidents to raise awareness [14]. While many studies have focused on detecting unsafe content on social media, there remains a significant gap in research regarding the types of safety topics that are shared online and the nature of the comments associated with such posts. Most research has focused on identifying harmful behaviors like

harassment but less attention has been given to understanding the broader conversation surrounding safety-related content itself. Specifically, there is limited exploration of how safety topics—such as safety practices, or responses to incidents like harassment—are framed and discussed by users. Additionally, the way in which comments and responses to these posts shape public perceptions of safety and influence the broader narrative around responses to incidents is underexplored. Filling this gap could provide valuable insights into how narratives of safety evolve and how social media platforms can better foster a supportive, informed, and safe online environment. In this research, we use Reddit as an example of social media because it fosters the creation of subcommunities, each centered around specific topics and interests. These subcommunities, or subreddits, provide a unique environment to explore how safety-related content is shared and discussed. By focusing on these subreddits, we examine the types of safety topics that are highlighted, as well as how users engage with such content through their comments. Through content analysis, we study user behaviour, analysing the patterns of interaction in response to safety posts.

2. Literature review

Safety conversations on social media

Users of social media do not shy away from discussing hot topics, whether politics, religion, and safety related topics. Disaster management have been well studied on social media. A systematic literature review analysed more than 100 relevant research papers that studies terrorism topics on social media and how this content was used to leverage a safer community [15]. There is a number of research that used topic modelling to uncover conversations of safety on social media. In one of the early works in this field and a highly cited work by Chen, Zhou, Zhu, and Xu (2012), the authors used text mining on social media to identify derogatory terms and profanity in an effort to enhance teen safety on social media [3]. More recently, in disaster management, social media users were found to discuss three main topics of health care and accidents namely: enhancing dealing with medical accidents, improving safety management, and media dealing with medical accidents [16]. In a similar realm, a study looked at COVID19 sentiment and main topics shared on social media and found that fear is the highest shared emotion in social media [17].

Moreover, researchers studied communication during a man-made disaster and the topics discussed on social media by mining online content and using topic modelling approaches [18].

Ali et al. (2023) studied DM (direct messages) in Instagram and asked users to identify risky topics [19]. The study found that length which is how long the user was committed to responding and taking part in the conversation was a predictor for risky encounters on social media. Although this was not a topic of safety, but it is an indicator of an unsafe topic.

A number of safety topics discussed on social media is identified in the literature, where majority of those topics are identified using data mining and algorithmic classifications. These methods are prone to error and lack the depth of human analysis due to the complexity of natural language. Therefore, in this research we use qualitative content analysis to understand the following question: RQ1. What safety related conversations people discuss in community created platform such as Reddit?

Engagement through comments on social media

Commenting on social media has profound effects on users' behaviours, and those behaviours might cascade to more significant reactions. Many research studied comments effect on news engagement on Facebook. For example, a study compared between comment reading behaviour and its effects on news reading and found a difference between western and eastern users' reactions [20]. In a similar realm, a study investigated comments on news posts and found that negative comments attracted users – using eye tracking – and conversely, had an effect on their sharing behaviour [21]. In a larger effect for comments, social media firestorms are an interesting result of the comments feature. Users would hail a large number of comments similar to having a heated discussion to either support or oppose an idea. Research studies the reasons behind engaging in a SMF and found that preconceptions in the real world would cascade online and would influence users' comments [22]. It was also found that such efforts are not always destructive as they might serve a higher purpose such as raising awareness or increasing public good. Similarly, although negative comments can have a negative effect on users' self-confidence, online support communities can alleviate some of those anticipated negative effects [23].

A systematic literature review examined papers discussing comments. The authors found that only five papers studied comments on Reddit, and that the number of studies focusing on comments understanding is decreasing, despite the increase in users of social media globally [24]. Interestingly, the authors emphasize the lack of papers discussing comments in the context of governments and local municipalities and local enforcement, therefore, signifying a need to understand comments in the

context of safety [24]. Therefore, we investigate the following question:
RQ2. What is the comments' categorization of users engaging in safety related content on Reddit?

3. Methodology

Data Collection
The data for this study were collected from Reddit, a popular social media platform known for its diverse and user-driven discussions. Specifically, we focused on subreddits for safety and posts related to the topic of safety. These subreddits discussed content in the U.S., therefore, the sample is from the U.S. Over a two-month period, the author manually retrieved 70 Reddit posts and 206 associated comments across various safety-related subreddits. The selection criteria were based on relevance to the central theme of safety, encompassing personal, community, and public safety topics. Posts were filtered to exclude irrelevant discussions, advertisements, and low-quality content (e.g., posts with less than two comments).

Content Analysis
A qualitative content analysis was employed to explore recurring themes, patterns, and sentiments related to safety concerns expressed on Reddit. Content analysis allows for the systematic categorization of textual data to identify meaningful patterns [25]. The coding was performed by three independent coders, including the primary researcher and two additional coders trained in qualitative analysis. The first coder independently reviewed and coded all 70 posts and 206 comments. Then, the three coders met weekly to review and comment on the coding process and to discuss the derived categories. Interrater reliability was measured using Percent Agreement. Although it is a very simple technique, previous research looked at 7 different indices including Cohen's kappa and found that the 7 Percentage agreement performed comparingly well [26]. Regular discussions were held to resolve discrepancies and refine the coding framework, ensuring that the final coding reflected a shared understanding of the safety themes present in the data. The overall percentage Agreement was 80%.
Coding discrepancies were discussed collaboratively, and adjustments were made to the coding scheme to enhance clarity and consistency. In every meeting, the coders reviewed the interpretations to validate the final thematic categorizations. This process ensured the robustness of the content analysis, minimizing subjective bias and improving the overall validity of the findings.

4. Results

To answer the first research question (what safety related conversations people discuss in community created platforms such as Reddit?) we found the following themes and subcategories:

Spreading awareness
The first high level category found was spreading awareness. Many users of the social media platform posted threads about raising knowledge about safety in general. This was witnessed in three different categories namely: sharing prior experience, posting safety regulations, and posting news about safety.
- Sharing prior experience:
In this category, users shared advice based on their previous experiences. For example, one user "Megan" posted about her experience receiving calls to relief people from the California fire. She said that a lady who was rescued had her rescue call and ask to get her medications that she lost in the fire. Unfortunately, she had a seizure, and Megan wasn't able to help her. She posted about the incident with a recommendation: "All I can think to do to ease my mind is to post this here. It is a link to a PDF that you can put all of your medication information on and keep in a couple places. My grandpa who was diabetic had it in his car in the medicine cabinet and in his wallet. You do not know when a disaster is going to strike. A lot of people who ran from this fire did not have the time to find their wallets or prescriptions.
Another example that is less severe is shared by Sally. She is addressing a physical safety injury: "Safety 1st always use the wrist strap kids! Trust me I've broke my TV on Wii sports before. Similarly, John shared his experience in working with metal and not wearing proper protection. He said: "I recently got metal in my eyes for not using eye protection... If only I thought about safety and used safety glasses. But I just use my regular glasses"
- Posting general safety tips
In this category, users have posted threads about safety tips they wanted to share with their community. Some posts included sources for their tips, for example: "…Information to help build your awareness in an active shooter event, and respond to options that have been proven to be most effective." However, majority of the posts did not include citations or verified websites to back up the shared information. For example:
"If someone points a gun at you and orders you to get into their vehicle, don't do it. You have a much, much better chance of surviving a GSW than you do of surviving whatever the suspect wants to do to you. If a cop tries to pull you over in a dark, secluded area, don't pull over. Obeying all traffic laws, keep driving

until you reach a well-lit, populated area. As you drive, call 911 and explain the situation. They will be able to verify that it's a real cop behind you, and will also be able to communicate to officers that you aren't trying to run on them. Any half-decent cop will understand." While this might seem like good advice, it is based on assumptions such as being a "half-decent" cop, and assuming the cop will understand the lack of compliance.

Similarly, a user posted a link to an image of general safety tips. The user put a heart and a high-five emojis next to the unvetted image.

-Posting news about safety

Reddit is a news sharing and commenting social media app, therefore, it is expected to see news posts related to safety in this app. All the posts in this category included a link to an external news article related to safety in general. For example, one user posted about a new school safety package. Other posts were about introducing a new app for transportation safety and a product that can improve safety against sexual assaults. It is worth noting that none of those news outlets shared was from a top news channel or from a government website, therefore, requires further vetting and verifications.

Seeking Advice

This category is the opposite of the previous one. Here, the users would post threads to ask about certain things and get safety related advice from other content generators. Looking at the collected posts, we found two main sub-categories, users post about an actual situation where they request advice, and users posting about a hypothetical situation and discuss outcome possibilities.

A. Actual scenario

In this category, users discuss a real situation where they need advice. In the dataset obtained, the researchers found that users discuss actual scenarios related to: personal, professional, and community related safety dilemmas.

The largest category discussed was personal, where users had questions related to their pets' safety, travel safety, food safety, and health safety. For example one post read: "How safe is Paris. I don't know if this falls under this category but my friend really wants to go [to] France but I don't know any safe places in France I've heard a lot that France is unsafe how true is this?". Similarly, a post about food safety reads: "What are the probabilities that I'll get sick from eating an uncooked pizza that was left out overnight, refrigerated, then cooked?". Also, kids safety was discussed in the thread. For example, a user posted: "Hi, I have little kids- 3 and 5- and am staying in a house with a pool during an upcoming vacation. There is no fence around the pool so my anxiety is

sky high. I started researching pool alarms. Has anyone here had experience with the safety turtle wrist bands?".

Actual scenarios shared where sometimes related to professional conduct. the situations were almost dichotomous, having either a single affect, or a cascading effect on others. An example of single effect posts is an employee asking about safety shoes in a plant. Another example is this post: "Am I safe from radiation/microwaves while working within a couple hundred feet of this tower for months at a time?". Here, the user was concerned for their safety only and asked a question for themselves. Other posts discussed larger safety effects such as: "Requesting help to find a Construction Site Safety inspection checklist … Really not finding much from OSHA nor MSHA available for download. Is this something that must be created in house, or is there a template floating around?". Similarly, a user posted: "Any Tips on Creating a WHMIS Training Program?". Sometimes posts are related to community safety where the user is concerned with the general public safety of their local community. For instance, a mother posted a picture of a playground and commented: "Does anyone else see a Safety Hazard here? Park staff not concerned. What should I do?"

B. Hypothetical scenario

Users on Reddit discussed some hypothetical situations and requested advice from other users. These posts were basically what-if questions that imagine a situation and start a conversation around it. A user posted this question "How would I talk to 911 if someone where actively chasing me around my house?" sparking conversation about tactics and tips from commenters who are not law enforcement. Some questions were about taking precautious actions and being proactive instead of reactive. For example, a user posted: "How do you survive a stampede? What should you do when you realize it's about to happen?". All posts in these categories seek safety advice from unprofessional and untrained people on social media platforms.

Seeking support e.g. asking if people need a free software or tool to improve safety in workplace.

Seeking support overlaps with seeking advice in the sense that users are requesting help from other users. However, the main difference is the nature of the request, where its to get written advice, or to request an action from the online community. In this section, we present the findings related to user-generated posts seeking support in three primary areas: petition support, job requests, and job search. Our analysis reveals the prevalence and nature of these requests, highlighting the motivations behind users' outreach

and the communal responses they elicit. For example, a user was working on a safety-related petition and asking for community support: "Hi there - petition for technology for children's safety. Please sign and share". Some users were seeking support in finding companies that can assist in enforcing safety. For example, a user posted: "I'm looking for a company that makes short safety videos, 5 mins or below.". on the other hand, some users were looking for jobs in the field of safety: "Anyone know a good subreddit for companies looking to hire ? I'm leaving the Military with 8 years Health and Safety experience, only an associates degree in the field." Similarly, a user posted: "Is anyone working in the midwest who would be willing to let me job shadow them? I am 23, college grad with a BS in Biology. I am trying to figure out what career I want to get serious about and would like to learn more about Safety". Moreover, there were some posts about general support such as an employee who is forced to work in a chemical lab with no safety training and is requesting community support to deal with this issue.

By categorizing the types of support sought, we aim to understand the dynamics of user interactions and the collective knowledge-sharing that occurs within these online communities. This examination provides valuable insights into how individuals navigate their needs for assistance and solidarity in professional and civic contexts.

Offering support

Very few posts offered support in safety matters. Majority of posts related to safety were triggered from a need. Therefore, in this category, only few posts offered safety support without a request. A user posted: "Would anyone benefit from some free software tools to help them manage their Environmental Health & Safety at work? Would you benefit from a free Incident Investigation tool/workflow or maybe a Job Site Safety Plan? What tools can I build you, a safety professional, that would make your job easier? What can Safety Management Systems do better?"

Safety related jokes (humor) usually include media i.e. picture or video

The last found category was related to joking about unsafe situations. The majority of posts in this category were in the form of an image, a meme, or a short video. These included jokes about previous safety incidents, safety advice in the form of a funny comic, or dark humor—just funny jokes about dangerous or risky scenarios. These posts did not always directly contribute to safety awareness but instead used humor to lighten the conversation around serious topics. While they might not always

offer actionable safety tips, they served to engage users by blending humor with awareness. This type of content often relied on exaggerated or absurd portrayals of safety issues, capturing attention in a more informal and accessible way. However, the line between humor and sensitivity could sometimes be blurred, raising questions about the potential impact on community perceptions of safety and the appropriateness of such humor in these discussions.

To answer the second research question (what is the comments categorization of users engaging in safety-related content on Reddit?), we found the following themes and subcategories:Based on the content analysis, we identified seven distinct categories for the comments of users engaging in safety-related content on Reddit. These categories help capture the diverse ways users contribute to discussions surrounding safety topics. Table 1 depicts the mapping of the results of research question 1 and research question 2.

Guidance

The first category we identified in our analysis is Guidance, which encompasses comments where users provide explanations, advice, or recommendations to help others navigate safety-related topics or challenges. These comments are typically informed by personal experience, expertise, or practical knowledge, and are aimed at offering clarity and direction to fellow Redditors. An example of this category can be seen in a comment discussing safety concerns related to antenna systems. One user provides a technical explanation, stating: "The dish antennas are very directional and very high and wouldn't be a problem unless they were high power (they are too small for that). There would not be any radiation going downwards.

Table 1. Mapping results of safety topic with comment categorization

Safety topics / Comments categorization	I. Spreading awareness	II. Seeking advice					III. Seeking support	VI. Offering support	V. Jokes	Total comments	Percentage
		A. Actual scenario				B. Hypothetical					
		A.1.Personal	A.2.Professional		A.3.Community						
			Single effect	Multiple effect							
Guidance	5	5	13	8	1	7	29			**68**	33.0
Reassurance of worth			2	1			3			**6**	2.9
Social integration	24	8	10	7	1	4	15	1	8	**78**	37.9
Attachment					1	1				**2**	1.0
Nurturance		10	3	2		6	3	1		**25**	12.1
Reliable alliance	1	1	3	10		1	9	2		**27**	13.1
Total comments	**30**	**24**	**31**	**28**	**3**	**19**	**59**	**4**	**8**	**206**	**100.0**

The vertical antennas (probably mobile telephony) do radiate a bit downwards but you would barely notice them even if you were sitting next to them on the tower."

This comment serves to clarify misconceptions about radiation from antennas, offering a helpful and detailed explanation that reduces fear and encourages understanding.

Another example of the Guidance category can be found in comments related to career advice in safety fields. One Redditor, when asked about subreddits focusing on companies hiring for safety roles, responds: "I'm not familiar with any subreddits that focus on companies hiring for safety. Check the usual job sites like Indeed and join your local ASSP and network is the best advice I can give." This response offers practical advice for those seeking safety-related jobs, recommending trusted platforms and professional networks that can help individuals connect with potential employers.

These examples illustrate how the Guidance category includes comments where users share knowledge or practical steps, whether it's explaining a technical issue or offering actionable career advice. The focus is always on helping others by providing valuable insights to navigate safety-related challenges effectively.

Reassurance of worth

The second category we identified in our analysis is Reassurance of Worth, which consists of comments where users offer encouragement, affirmations, or validation to others regarding their safety knowledge, skills, or career prospects. These comments often serve to bolster confidence and reassure individuals that their efforts or qualifications are valuable, especially when they may be uncertain about their abilities or the next steps in a safety-related journey.

An example of this category can be seen in a comment offering career advice: "… your resume will easily get you an entry/mid-level position with wing safety anywhere there would be an opening." This comment reassures the individual that their qualifications are likely to be sufficient to secure a position in the field, providing a sense of confidence that their efforts are worthwhile. By validating the potential for success, the commenter helps to alleviate any doubts about entering the safety profession. Similarly, another comment in this category emphasizes the importance of self-development in the field: "I highly advise developing your own, with the help of a safety manager or one of the many consulting agencies." This statement encourages users to take ownership of their growth and development, assuring them that seeking guidance from professionals can lead to personal

success in safety-related work. It validates the notion that building a personal safety strategy or plan with professional assistance is both feasible and a smart approach.

Social integration

The third category identified in our analysis is Social Integration, which includes comments that foster a sense of belonging and connection within the safety-related community or broader social context. These comments often serve to establish rapport, offer shared experiences, and help individuals feel included in discussions or activities. They are characterized by a friendly tone, emotional support, and a sense of camaraderie.

An example of Social Integration can be found in a comment discussing personal experiences with safety training: "I hope your TAPS instructor was as good as mine, I landed quite nicely once the 214 was official." Here, the user shares a positive personal experience and encourages connection through a common training program. The reference to the instructor and the successful outcome helps establish a bond with others who may have undergone similar training, reinforcing a sense of shared experience and mutual understanding within the safety community.

Another example in this category is a more casual and lighthearted piece of advice: "Seriously, go have fun, channel your inner shoe loving niece, have a nice time trying everything on. It's your feet, you only have two, and you'll likely be wearing these shoes a lot!" In this case, the comment encourages users to enjoy the process of selecting safety shoes, while also acknowledging their importance for comfort and long-term wear. The friendly and relatable tone helps users feel more comfortable and connected in making decisions about their safety gear, fostering a sense of community and shared experience in a typically practical and sometimes overlooked aspect of safety

Nurturance

Nurturance includes comments that provide emotional support, care, and encouragement to others, especially in situations where individuals may feel uncertain, vulnerable, or anxious. These comments often aim to ease concerns and offer guidance with a nurturing and compassionate tone, helping others feel supported and reassured.

An example of Nurturance can be seen in a comment offering safety advice related to exposure to the sun: "Nothing on the tower will harm you if you're outdoors either, but the big ball of radiation in the sky will give you a bad time. Wear sunscreen and cover up." This comment provides not only practical safety advice but also a sense of concern for the well-being of others. Another example of Nurturance

comes in the form of advice regarding workplace dynamics: "Don't go to HR. They are not your friend. They are there solely to protect your company, and do not have your best interests in mind." Here, the user offers a form of emotional protection, advising the individual to consider alternative approaches when dealing with workplace issues. This comment aims to shield the user from potential workplace challenges by giving advice that might mitigate negative experiences.

Additionally, a more reassuring advice: "You're fine, worst case an extra trip to the bathroom." This comment provides emotional comfort in a potentially embarrassing or uncomfortable situation, easing the person's worries and reassuring them that the outcome is not as bad as they might think. By downplaying the potential problem and offering a calming perspective, the user provides a sense of relief and emotional support.

Reliable alliance

Reliable Alliance, which, while overlapping with some of the previous categories, is distinct in that it involves conversations between professionals offering specialized, profession-related support. These comments are focused on providing practical, safety-focused advice that reflects expertise and a shared commitment to best practices within the safety field, rather than just emotional or general support.

For example, one user asks for professional advice about workplace safety: "I'm not trained to work with asbestos... that's the basis on which I refused to work with it. Would it be an acceptable HR complaint if my manager tried to coerce me into doing it without training?" This comment reflects a professional seeking guidance on navigating safety responsibilities and company protocols, emphasizing the importance of relying on professional standards in the workplace. Similarly, another user emphasizes the significance of proper qualifications in handling hazardous materials: "Having unqualified employees do this is asking for an accident. We collect cylinders, but we don't purge them. If you need better advice, contact your local occupational health and safety representative." This reflects professional responsibility in ensuring safety and encourages others to seek out specialized knowledge when handling dangerous tasks. Also, a user offers helpful resources with a professional touch: "I have several great templates... Free of course, message me and I'll send them when I get to the office Monday." By offering approved templates for safety procedures, the commenter provides profession-specific resources to help others adhere to safety standards. These examples illustrate how the Reliable Alliance category is characterized by exchanges of profession-specific advice and support, where users share specialized knowledge and resources to ensure safety. Unlike emotional or general support found in other categories, this category centers around providing guidance and assistance that reflects a shared professional responsibility and expertise.

5. Discussion

To understand the results and interpret them under two magnifying theories, uses and gratifications and social support theory. Uses and Gratifications Theory (UGT) coined in 1973 by Katz, Blumler, and Gurevitch posits that users purposefully use media to receive a form of gratification and/or serve a pressing need [27]. Since the study investigates comments and engagement, we also use the social support theory as a lens to understand the results. This theory discusses how social support is needed from friends and relatives to fulfil a need and provide support and information when needed [28].

The results of the study on Reddit users engaging in conversations related to safety reveal a dynamic interplay of support-seeking and support-giving, which can be understood through the lens of Social Support Theory. This theory emphasizes how individuals use their social networks to seek emotional, informational, and instrumental support in times of need. The findings of this study show that Reddit users engage in safety related conversations where they both seek and provide various types of support, drawing from past experiences as well as hypothetical scenarios. The study found that many Reddit users turn to the platform to seek emotional and informational support related to safety concerns. Whether discussing personal safety, home security, or general advice on feeling secure in different environments, users often post about their worries and concerns. This mirrors the emotional support dimension of Social Support Theory [28], where individuals seek comfort, reassurance, and validation from others to alleviate stress and anxiety. Additionally, users often ask for informational support, such as advice on safety measures, recommendations for self-defense tools, or tips on securing their online presence. This aligns with the informational support aspect of the theory, where individuals rely on others to provide guidance or feedback to help them make better decisions. However, providing unvetted support is dangerous and is proven to adversely affect people, specifically users with lower education [29].

The study also found that both past experience and hypothetical situations play a key role in how users engage in conversations on safety. Many Reddit users share personal stories of incidents they have faced, which serve as real-life examples to provide

reassurance or caution to others. This use of past experience, which is biased by nature, offers well-intentioned support, as users evaluate and share the outcomes of their actions to help others assess their own safety strategies. On the other hand, discussions involving hypothetical situations allow users to explore potential safety issues that have not yet occurred but may be of concern. This speculative approach can also be helpful in fostering preparedness and problem-solving. By using hypothetical situations, users engage in a form of informational support that encourages proactive thinking and planning for safety challenges. These findings highlight the importance of social networks in shaping how individuals cope with safety-related concerns. Reddit, as a platform, serves as a virtual community where users engage in reciprocal support exchanges, providing both emotional comfort and practical advice to one another. Through this interactive process, individuals can manage their safety concerns more effectively, relying on the collective knowledge and experiences of others. In conclusion, the results of this study highlight how Reddit users actively engage in conversations about safety that reflect the key components of Social Support Theory. By seeking and offering various types of support—emotional, informational, and instrumental—users create a supportive environment where both personal experiences and hypothetical scenarios contribute to greater safety awareness and problem-solving. This exchange of support not only helps individuals feel more secure but also strengthens the sense of community among users, which is central to the theory's focus on the role of social connections in well-being.

While social support theory helps us understand the nature of seeking socialistic behaviours and responses, Uses and Gratifications Theory allow us to explain the results of users who discuss safety topics to receive a sense of gratification or reward [27]. This theory focuses on how individuals actively seek out media and communication platforms to fulfill specific needs and desires. In the context of the study, Reddit users engage in safety discussions to satisfy various personal needs, such as seeking emotional comfort, obtaining practical advice, or fulfilling a sense of belonging within a community. The act of posting questions, sharing personal experiences, or providing feedback on safety matters reflects users' active role in using the platform to fulfill those needs. This dynamic aligns with the core idea of Uses and Gratifications Theory, where individuals are seen as active participants, rather than passive receivers, of information.

Users seeking emotional support in safety discussions are likely looking for reassurance or validation that their concerns are legitimate, which

speaks to the emotional gratification aspect of the theory. By engaging with others who share similar experiences, they can alleviate anxiety and gain comfort. Meanwhile, those seeking informational support are using Reddit to fulfill a practical need: advice on how to secure their homes, protect themselves, or mitigate potential risks. This aligns with the theory's idea that people use media to satisfy specific informational needs, seeking resources that help them make better-informed decisions.

Social Support Theory complements and deepens the understanding of the results by providing a more nuanced framework for analyzing the type of support that is exchanged. While Uses and Gratifications Theory emphasizes why people turn to platforms like Reddit—primarily to fulfill needs related to safety—Social Support Theory explains how the interactions unfold in terms of the types of support exchanged. Social Support Theory categorizes support into emotional, informational, and instrumental types, which are all evident in the study. Emotional support manifests as users offering empathy and reassurance, informational support comes through practical advice and knowledge sharing, and instrumental support appears when users suggest concrete safety measures or tools.

By looking at results under both theories, we can see a more nuanced picture of why and how individuals engage in conversations about safety on platforms like Reddit. Uses and Gratifications Theory helps us understand the motivation behind these interactions—how users actively seek specific types of support—while Social Support Theory explains the types of support exchanged and the social dynamics that influence these exchanges.

Design Implications

Crowd-sourced verification is a powerful method for improving the accuracy and trustworthiness of safety-related information on social media platforms. It leverages the collective intelligence of users to assess and validate the credibility of content. This approach involves allowing the community to act as an additional layer of content moderation and fact-checking, helping to quickly identify misinformation or unreliable safety advice. Key features of this system include voting systems where users can rate or flag content as accurate or inaccurate, transparency in feedback so users can see how information is rated, and badges or verification statuses for trusted individuals or experts. These elements create a dynamic and interactive system where users can collaboratively ensure that shared safety information is reliable and trustworthy.

In practice, crowd-sourced verification methods can take various forms. One method involves user voting, where individuals rate the accuracy of a post

on a scale or through binary voting (e.g., "Yes" or "No" on accuracy). Another approach is to involve experts or authority figures who contribute verified information, providing users with authoritative responses to safety-related questions. Platforms can also partner with third-party fact-checking organizations to provide an additional layer of validation. Peer review mechanisms, where multiple knowledgeable users assess the content, can also contribute to ensuring information is correct and reliable. By tapping into a broad user base, this approach helps distribute the responsibility of content moderation, making it scalable as social media platforms grow.

The advantages of crowd-sourced verification are numerous. It increases the accuracy of safety-related discussions by tapping into the collective knowledge of the user base. This system also scales well, allowing platforms to manage large volumes of content without relying solely on human moderators. Furthermore, it democratizes content moderation, giving users more control over the information they encounter and share. This can foster greater trust among users, especially when safety-related misinformation could have serious consequences. However, there are challenges, such as the potential for misinformation to spread within the crowd, difficulty in coordinating large numbers of users, and the risk of system manipulation by bad actors. To mitigate these risks, platforms must introduce safeguards, such as reputation systems to reward users who consistently contribute accurate information and limit the influence of those who attempt to game the system.

To implement an effective crowd-sourced verification system, social media platforms can create tiered user levels, with users earning reputation points for their contributions. Verified experts or users with higher reputation scores could have more influence in the verification process. Additionally, transparency is crucial—users should be informed when content is flagged or voted down and provided with feedback to improve future contributions. Real-time alerts and updates could notify users when a post they interacted with has been verified or flagged, keeping them informed about the accuracy of the information they engage with. An example of this could be seen in health or safety-related topics, where crowd-sourced verification could help ensure that posts about emergency preparedness or mental health resources are based on current, accurate information. Overall, crowd-sourced verification can enhance the safety of online spaces by allowing the community to actively participate in ensuring the accuracy of safety-related content, but it must be carefully designed and managed to avoid misuse and maintain trust.

Encouraging responsible sharing of information is essential for maintaining safety and trust on social media platforms, especially when it comes to sensitive topics such as personal safety, health, and emergency situations. Many users may unknowingly share misinformation or advice that may be outdated, misleading, or even harmful, simply because they lack awareness of its potential consequences. To address this, platforms can implement design features that promote more thoughtful and informed sharing. For example, before posting or sharing safety-related information, users could be prompted to check the accuracy of their content by offering easy-to-use fact-checking tools or displaying reminders about the importance of sharing reliable information. Additionally, platforms can encourage users to provide sources or cite evidence when sharing safety tips or warnings. This could be implemented through user interface elements such as a simple prompt or checkbox that reminds users to include credible sources when they post about safety or public health topics.

Another effective strategy is to integrate systems that offer feedback on the quality of shared content. This could involve flagging posts that are highly likely to be inaccurate or misleading based on community feedback or automated detection of commonly known misinformation. When users see that posts they interact with have been flagged or reviewed, it may encourage them to think more critically about the information they are consuming and sharing. Moreover, the platform could provide users with brief tutorials or educational pop-ups that explain how to recognize and avoid misinformation, helping them become more responsible content creators.

Additionally, social media platforms could introduce a system where users are rewarded for consistently sharing accurate, well-sourced information. This can be paired with visible indicators that show which users are known for sharing high-quality, trusted information, providing a form of recognition that encourages others to follow suit. Such mechanisms not only incentivize responsible sharing but also help create a network of users who can be relied upon to disseminate trustworthy information. Ultimately, by fostering a culture of responsibility around information sharing, platforms can help reduce the spread of harmful misinformation and ensure that safety-related conversations contribute positively to the well-being of users.

Limitations

A limitation of this research is the manual data extraction from Reddit and the manual coding process, which may introduce bias due to the subjective interpretation of data. Additionally, focusing on only one social media platform limits the

generalizability of the findings, as safety conversations may vary across different platforms. Comparing these results with data from other social media platforms would provide a more comprehensive understanding of safety-related discussions and reduce potential platform-specific biases.

6. Conclusion

This study highlights the significant role that online communities, particularly on platforms like Reddit, play in discussions about safety. Through content analysis, key themes and user interactions related to safety concerns were identified, showing how individuals turn to digital spaces for both informational and emotional support. The application of Social Support Theory and Uses and Gratifications Theory has provided valuable insights into the motivations behind users' engagement with safety content, ranging from seeking social connection to fulfilling practical needs. The findings demonstrate the importance of creating spaces that not only address immediate safety concerns but also foster a supportive environment for users to share experiences and resources.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1]Barak, A., Boniel-Nissim, M., & Suler, J. (2008). Fostering empowerment in online support groups. Computers in Human Behavior, 24(5), 1867-1883.

[2]Cobb, S. (1976). Social support as a moderator of life stress. Psychosomatic Medicine, 38(5), 300-314.

[3]Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. Psychological Bulletin, 98(2), 310-357.

[4]Katz, E., Blumler, J. G., & Gurevitch, M. (1973). Uses and gratifications research. Public Opinion Quarterly, 37(4), 509-523.

[5]Rogers, E. M. (2003). Diffusion of Innovations (5th ed.). Free Press.

[6]Ruggiero, T. E. (2000). Uses and gratifications theory in the 21st century. Mass Communication & Society, 3(1), 3-37

[7]Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159-174.

[8]Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. Qualitative Health Research, 15(9), 1277-1288

[9]Li, W., Wong, S., & Yu, P. (2019). Interrater reliability estimators tested against true interrater reliabilities. BMC Medical Research Methodology, 19(1), 168. https://doi.org/10.1186/s12874-019-0700-0

[10] Team, B. (2025, February 10). Social Media Usage & Growth Statistics. Backlinko. https://backlinko.com/social-media-users

[11] Team, B. (2025a, January 21). Reddit user and growth stats. Backlinko. https://backlinko.com/reddit-users

[12] Redmiles, E. M., Bodford, J., & Blackwell, L. (2019, July). "I just want to feel safe": A Diary Study of Safety Perceptions on Social Media. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 13, pp. 405-416).

[13] Ali, S. (2024). Youth, social media, and online safety: a holistic approach towards detecting and mitigating risks in online conversations (Doctoral dissertation).

[14] Gibson, A. (2019). Free speech and safe spaces: How moderation policies shape online discussion spaces. Social Media+ Society, 5(1), 2056305119832588.

[15] Chen, Y., Zhou, Y., Zhu, S., & Xu, H. (2012, September). Detecting offensive language in social media to protect adolescent online safety. In 2012 international conference on privacy, security, risk and trust and 2012 international confernece on social computing (pp. 71-80). IEEE.

[16] Chowdhury, A. G., Sawhney, R., Shah, R., & Mahata, D. (2019, July). YouToo? detection of personal recollections of sexual harassment on social media. In Proceedings of the 57th Annual

Meeting of the Association for Computational Linguistics (pp. 2527-2537).

[17] Abarna, S., Sheeba, J. I., Jayasrilakshmi, S., & Devaneyan, S. P. (2022). Identification of cyber harassment and intention of target users on social media platforms. Engineering applications of artificial intelligence, 115, 105283.

[18] Chowdhury, A. G., Sawhney, R., Mathur, P., Mahata, D., & Shah, R. R. (2019, June). Speak up, fight back! detection of social media disclosures of sexual harassment. In Proceedings of the 2019 conference of the North American chapter of the Association for Computational Linguistics: Student research workshop (pp. 136-146).

[19] Cai, J., Wohn, D. Y., & Almoqbel, M. (2021, June). Moderation visibility: Mapping the strategies of volunteer moderators in live streaming micro communities. In Proceedings of the 2021 ACM International Conference on Interactive Media Experiences (pp. 61-72).

[20] Almoqbel, M., Wang, X., & Hiltz, S. R. (2020). Do I care enough? using a prosocial tendencies measure to understand Twitter users sharing behavior for minor public safety incidents.

[21] Almoqbel, M., & Xu, S. (2019). Computational mining of social media to curb terrorism. ACM Computing Surveys (CSUR), 52(5), 1-25.

[22] Kim, N., & Lee, N. J. (2024). Structural Topic Modeling Analysis of Patient Safety Interest among Health Consumers in Social Media. Journal of Korean Academy of Nursing, 54(2), 266-278.

[23] Monselise, M., Chang, C. H., Ferreira, G., Yang, R., & Yang, C. C. (2021). Topics and sentiments of public concerns regarding COVID-19 vaccines: Social media trend analysis. Journal of medical Internet research, 23(10), e30765.

[24] Deng, Q., Gao, Y., Wang, C., & Zhang, H. (2020). Detecting information requirements for crisis communication from social media data: An interactive topic modeling approach. International Journal of Disaster Risk Reduction, 50, 101692.

[25] Chen, Y., Zhou, Y., Zhu, S., & Xu, H. (2012, September). Detecting offensive language in social media to protect adolescent online safety. In 2012 international conference on privacy, security, risk and trust and 2012 international conference on social computing (pp. 71-80). IEEE.

[26] Ali, S., Razi, A., Kim, S., Alsoubai, A., Ling, C., De Choudhury, M., ... & Stringhini, G. (2023). Getting meta: A multimodal approach for detecting unsafe conversations within

instagram direct messages of youth. Proceedings of the ACM on Human-Computer Interaction, 7(CSCW1), 1-30.

[27] Matook, S., Dennis, A. R., & Wang, Y. M. (2022). User comments in social media firestorms: A mixed-method study of purpose, tone, and motivation. Journal of Management Information Systems, 39(3), 673-705.

[28] Alafwan, B., Siallagan, M., & Putro, U. S. (2023). Comments Analysis on Social Media: A Review. EAI Endorsed Transactions on Scalable Information Systems, 10(6).

[29] Almoqbel, M. Y., Wohn, D. Y., Hayes, R. A., & Cha, M. (2019). Understanding Facebook news post comment reading and reacting behavior through political extremism and cultural orientation. Computers in Human Behavior, 100, 118-126.

[30] Dutceac Segesten, A., Bossetta, M., Holmberg, N., & Niehorster, D. (2022). The cueing power of comments on social media: how disagreement in Facebook comments affects user engagement with news. Information, Communication & Society, 25(8), 1115-1134.

[31] Ismail, L., Fatmawati, F., & Nursida, A. (2024). The Impact of Hate Comments on Social Media Users' Self-Confidence: An Analysis Grounded in Social Comparison Theory and Cyberbullying Research. Acta Psychologia, 3(2), 49-59.

[32] Katz, E., Blumler, J. G., & Gurevitch, M. (1973). Uses and gratifications research. The public opinion quarterly, 37(4), 509-523.

[33] Higgins, P. M., & Kurtzman, J. G. (1988). Guy M. Rogers, Wellesley College, MA, The History-Non-US. Humanities, 1956, 37.

[34] Oviedo-Trespalacios, O., Peden, A. E., Cole-Hunter, T., Costantini, A., Haghani, M., Rod, J. E., ... & Reniers, G. (2023). The risks of using ChatGPT to obtain common safety-related information and advice. Safety science, 167, 106244.

[35] Shohan, M. H., Ahmed, K. R., Kahar, N. F., Jahan, N., & Maruf, M. (2025). Use of Natural Language Processing for the Detection of Hate Speech on Social Media. Journal of Advanced Research in Applied Sciences and Engineering Technology (ARASET), 51(2), 86-96.

[36] Achara, A., & Chhabra, A. (2025). Watching the AI Watchdogs: A Fairness and Robustness Analysis of AI Safety Moderation Classifiers. arXiv preprint arXiv:2501.13302.

[37] Mitts, T. (2025). Safe havens for hate: The challenge of moderating online extremism. Princeton University Press.