Analyzing Reddit Stories of Sexual Violence: Incidents, Effects, and Requests for Advice

Vaibhav Garg, Hannah Javidi, Ruijie Xi, Jiaqing Yuan, Munindar P. Singh
Department of Computer Science, Virginia Tech, Alexandria, VA
Department of Applied Health Science, Indiana University, Bloomington, IN
Meta, WA
Amazon Web Services, NY
Department of Computer Science, North Carolina State University, Raleigh, NC
vaibhavg@vt.edu, hsjavidi@iu.edu, rxi@ncsu.edu, jordanyuan111@gmail.com, mpsingh@ncsu.edu

Abstract

Warning: This paper may contain triggering language for some readers, especially survivors of sexual violence.

Survivors of sexual violence sometimes share their experiences on social media, revealing their emotions and seeking advice. On platforms such as Reddit, some stories can be long—up to 40 000 characters. We posit that such long stories are demanding for helpers to read and respond to. Prior research has indicated that parts of these stories describing the incident, the effects on the poster, and advice requested by the poster are important. Highlighting those parts can draw helpers' attention toward key information and assist them in reading and responding to long stories.

We first examine the stories posted on Reddit for the prevalence of these parts. Second, we develop a computational model to highlight these parts of a story. On ten-fold cross-validation of a dataset, our model achieves a macro F1 score of 0.82. In addition, we contribute METHREE, a dataset comprising 8947 labeled sentences for these parts from Reddit stories.

A survey of users who are helpers on some relevant subreddits shows that the parts highlighted by our tool represent important information and assist them while reading and responding to long stories. We find that these tool-generated highlights statistically significantly reduce the demandingness of long stories. Moreover, almost all helpers felt that highlighted stories are helpful and easier to read, understand, and respond to than nonhighlighted ones. In particular, on a four-point Likert scale, there is about 0.7 point reduction in demandingness when stories were presented with highlights.

1 Introduction

Sexual violence refers to unwelcome or nonconsensual sexual advances, verbal or physical acts of a sexual nature, along with gender-based insults, invalidations, and assaults (Gartner and Sterzing 2016; EEOC 2023; RAINN 2023). In the United States, 81% of women and 43% of men have reported some form of sexual violence in their lifetime (NSVR 2023). Survivors of sexual violence often feel anxious, depressed, or go through post-traumatic stress disorder (Field-Springer et al. 2022; O'Neill 2018).

Tarana Burke, a civil rights activist, coined the phrase "me too" to help survivors know they are not alone. This led to the *MeToo* movement, which involved people disclosing their sexual violence stories on social media. Social media helps survivors remain anonymous and hence be unafraid of potential public shaming. Sharing such stories facilitates survivors in venting out or seeking support from *helpers*, generally fellow users on the same platform. Helpers can aid survivors in ways such as healing and reporting abuse (Andalibi et al. 2016; O'Neill 2018).

1.1 Background

We investigate sexual violence stories on social media with a view to facilitating *helpers* providing support to survivors. We focus on Reddit, a popular social media platform that includes forums called subreddits. Of these, r/meToo (Reddit 2023a), r/SexualHarassment (Reddit 2023c), and r/sexualassault (Reddit 2023b) are geared toward survivors sharing their stories. Example 1 shows one such story. For this paper, we paraphrased these stories to reduce the traceability of the author. This is because even though the stories are public, it may not be appropriate to draw attention to survivors' identifiable details.

Reddit allows long posts (of up to 40 000 characters). The high limit is desirable because it ensures survivors can write freely and share any information they wish to. Survivors avail of this high limit by posting long stories: the 4933 relevant stories we collected have a mean length of 1881 and a maximum length of 33 432 characters.

1.2 Problem setting

We analyze long stories from the helpers' perspective of providing support. We conducted a survey of helpers which revealed that long stories could be demanding for helpers to read and reply to (discussed in Section 3). Hence, we pose the following research question.

RQ$_{\text{demanding}}$: Can the demandingness of long stories be reduced by highlighting their key components?

Example 1 shows a story of approximately mean length in which the survivor describes their experience and needs support about how to handle it. We include the instances of much longer stories in the appendix (Example 3).

Helpers provide support by responding to stories. However, long stories, such as those found typically seen on Reddit, challenge helpers. Demanding stories may attract few or delayed responses. Thus, whereas we want survivors to feel free to post as long a story as they desire, we would like to assist helpers in reading and responding to such stories. We found that highlighting important parts of stories assists helpers in reading and responding to them. Accordingly, we include the following research question.

RQ$_{\textbf{highlight}}$: How can we automatically highlight key parts of violence stories?

Example 1: A sexual violence story.

Parts: (Incident), [Effects], <Advice requested>

In 20xx, I went out with a group, consisting of my girlfriend and her two guy friends. She ended up canceling on us last minute so it ended up just being me and these two guys. I had hung out with those guys before so I thought they were trustworthy. We were out in the city driving around and drinking in the car and we three got wasted.

(After pulling me over in a dark area, where some vehicles were parked, I found that one of them was trying to kiss me. Most of the incident was blurry to me but I remember that he pressurized me to make out with him. Honestly, I don't think that I gave my consent to him but I also don't remember asking him to stop. I have blurry memories–pieces and gaps of being laid down after getting off my clothes. I also have memories of having his friend riding on me.) I woke up the next day and found myself naked. I wore my dress and asked him to take me home. This was because I thought my family was probably looking for me at this point. I was not in the state of thinking but rather wanted to get home ASAP.

Once I got home my parents were in tears and questioned me if anything had happened, she asked if I was raped? I said no promptly. How was I sure that it was rape and how could I convey the same to my parents? (Once I came into my room, I saw my body and found bruises on my hands and cuts and scrapes on my legs.) I kept telling myself that since I was drunk, maybe I started having intercourse with the guy. [But I could also feel his friend in me.] He messaged me the next day to ask about my health and told me that he had done nothing wrong, which confused me. So I buried this incident and moved on. When I look back, I understand and realize that I was raped by both of them that night.

[I am still getting glimpses of the assault and I can't stop thinking about it.] Things are not yet clear to me. [I am getting anxious after all these glimpses.]

<But Why now? Why am I having flashbacks and what to do about it? Do you recommend me to go to a counselor?>

1.3 Contributions and Findings

We make two contributions. For RQ$_{\text{demanding}}$, we surveyed a sample of users who serve as helpers on violence-related subreddits. On average, these users have written between four and ten supportive responses on such subreddits. We showed them long stories with and without highlights, which they rated on a demandingness scale. To answer RQ$_{\text{highlight}}$, we applied active learning to curate MeThree, a dataset of 8947 sentences drawn from stories in r/meToo, r/sexualassault, and r/SexualHarassment and labeled Incident, Effects, Advice requested, or None. (We describe these categories in Section 2.) In addition, we created a computational model for highlighting these important parts of a story. We validated the effectiveness of our model through the same survey. A survey of 29 helpers showed that helpers find long stories demanding to read and respond to. On a four-point scale of demandingness, the survey respondents provided mean ratings of 3.10 for reading and 3.13 for responding.

The Mann-Whitney U-test shows that highlights significantly reduce the demandingness of reading and responding to long stories. Moreover, almost all respondents prefer the highlighted stories as they are easier to read, understand, and respond to than nonhighlighted stories. On average, respondents found highlighted stories about 0.7 less demanding (on a four-point Likert scale) than nonhighlighted ones.

Qualitative findings suggest multiple variations in the structure of violence stories. This complexity in the story structure makes automatic highlighting difficult yet potentially useful to helpers.

Organization Section 2 explains the three parts of stories considered important in literature and our qualitative analysis based on them. Section 3 describes a survey of Reddit helpers to understand their perspectives. Section 4 explains how we collected Reddit stories and applied active learning to curate MeThree dataset and build a computational model. Section 5 discusses the potential impact of our work, including how our model can be applied on multiple subreddits. Sections 6, 7, and 7.2 describe the related work, discussion, and ethical concerns, respectively. An appendix presents additional details.

2 Diving Deep into Sexual Violence Stories

Three parts of violence stories are considered important in the literature. To demonstrate these parts, we highlight them in Example 1. We describe these parts along with the supporting literature below.

Incident Text describing unwelcome sexual advances, sexual behavior, requests or forcing for sexual favors, verbal or physical acts of sexual nature, offensive jokes or remarks that are either sexual or based on someone's gender (EEOC 2023; RAINN 2023; Ghosh Chowdhury et al. 2019a; Karlekar and Bansal 2018).

Example 1's text highlighted as (this) describes inappropriate touching—the incident here.

Effects on the survivor Survivors describe how they are affected by revealing their emotions that arise during or after the incident (Field-Springer et al. 2022; O'Neill 2018). Examples of effects include the survivor feeling uncomfortable due to the abuser's actions, or being angry or upset due to the incident.

Example 1's text highlighted as [this] is where the survivor discusses their discomfort due to this incident.

Advice requested Text in which survivors seek advice from other users (Andalibi et al. 2016; Rickwood et al. 2005; O'Neill 2018). Some examples of advice requested include asking if their (i.e., the survivor's) reported experience constitutes harassment or assault, how to pursue a legal case, and how to confront the abuser.

Example 1's text highlighted as <this> indicates that the survivor is confused and asks if they are overthinking the incident. Even if they are confused, such situations are clearly about violence, according to the incident described. Hence, key parts of such stories must be brought to the helpers' attention.

We randomly sampled, 360 stories (from 4933 relevant stories, whose collection is described in Section 4.1) and investigated them for the prevalence of the above three parts. To determine the size of the random sample, we applied Cochran's formula (Cochran 1977). The formula (used with standard values of 95% confidence, 0.5 proportion size, and 0.05 margin of error) suggested to sample 356 stories from 4933 relevant ones. We rounded this off to selecting 360 stories for our manual analysis.

Based on our investigation of these 360 stories, Table 1 shows the mean length for each of their three parts (incident, effects, and advice requested). On average, the incident part takes 29.66% of story text, which is greater than for effects (17.05%) and advice (6.99%).

Besides story length, an additional challenge is complexity, which translates to how demanding a helper finds understanding and responding to a story. We validated the importance of these parts and the overall demandingness of stories through helpers on subreddits, as discussed in Section 3.

Parts of a story	Mean proportion of length (%)
Incident	29.66
Effects	17.05
Request for advice	6.99

Table 1: Distribution of a story's length across the three parts in a random sample of 360 stories.

We analyzed the structures of 360 randomly selected stories in terms of these three parts. Example 1 shows the canonical story structure is I-E-A: *incident* first (I), followed by *effects* (E), and finally seeking *advice* (A) to overcome the current situation. However, that's not the case always. Our qualitative analysis revealed multiple variations in the story structure. For brevity, we show only two of the several possible structure variations in Example 2. The first story starts directly with seeking advice, which may be the main reason for a survivor to share their experience. Afterward, the survivor discusses the incident and its effects, leading to the A-I-E structure. The second story goes back and forth between incident and effects and ends with seeking advice. We found that some sentences belong to more than one part, leading to more complex variations in story structure. These variations complicate automatic highlighting.

Moreover, we found cases in which no advice is sought in the story. This is because of two reasons: (i) advice is already sought in the title of the post, or (ii) the survivor is only venting their experience and not seeking advice. In the latter case, helpers provide sympathetic and supportive responses to help the survivor feel better. To provide such sympathetic responses, reading and understanding long stories remains important. Hence, these stories are relevant to our study.

Example 2: Variations in story structure.

Parts: (Incident), [Effects], <Advice requested>

Story exhibiting the A-I-E structure

<Through this post, I want to know one thing: was I raped?> ... (Once, he was pressurizing me to lift me up my clothes and show him my breast.) ... [I hate myself for letting it happen and for not saying a word against it] ...

Story exhibiting the I-E-I-E-A structure

(My middle school teacher groomed me.) ... [When he used to do things, I used to feel humiliated and horrible.] ... (At another timestamp, another married coworker was touchy and flirtatous with me.)... [Suddenly, my anxiety started taking over and I had to quit my job.]... <Eventually being a victim, I developed having strange fantasies, is it okay to have such fantasies in my mind? Or should I just flush them out somehow?...>

Ten of these 360 stories included TLDR (Too Long; Didn't Read), a brief summary of the story optionally provided by the poster. Nine of the TLDRs described the incident, three described the effects, whereas five described the advice, indicating the importance of these parts even from the survivors' perspective. However, stories that include TLDR text are rare (10 out of 360), motivating our work to highlight the key parts.

3 Empirical Study of Helpers

We conducted a survey of helpers on Reddit due to two objectives. First, we investigate $\text{RQ}_{\text{demanding}}$ by directly asking helpers to rate the demandingness of highlighted and unhighlighted stories. Second, we analyze the benefit of the highlights generated by our model (model details in Section 4). In turn, this relates to $\text{RQ}_{\text{highlight}}$ as the survey validates the importance of three parts being highlighted, along with the effectiveness of our model.

We designed the survey with the help of a psychologist, who is an expert in understanding sexual violence and sexual consent. Additionally, the psychologist assisted in analyzing the survey responses. Since our study is primarily based on Reddit, we decided to recruit respondents from sexual violence-related subreddits.

Many sexual violence-related subreddits (including r/SexualHarassment, r/sexualassault, and others such as r/afterthesilence and r/Molested) don't allow such postings for research surveys. To recruit suitable respondents, we were left with two subreddits, namely, r/metoo and r/rape.

Interested respondents were asked to sign a consent form detailing the nature of the survey. Our study posed minimal risk to respondents and was exempted by the Institutional Review Board (IRB) at our university.

We used a priori power analysis (one sample case) to decide the desired number of survey respondents. For this analysis, we used the standard values of variables: alpha of 0.05, a power level of 0.8, and an effect size of 0.6. The test showed that we needed 20 respondents. Based on the survey engagement on subreddits, we ended up recruiting 29 respondents as it passed the desired sample size and was also managed within our funding.

The initial survey questions were related to the demographics of the respondents and were optional. Out of thw 29 respondents, 28 chose to reveal their gender identities (19 males, 8 females, and 1 nonbinary, genderqueer, or genderfluid) and 27 revealed their age (ranging in 22–38).

On average, these 29 respondents of the survey have written between four and ten helpful responses on violence-related subreddits. Upon survey completion, each respondent received an Amazon gift card of $20. Respondents were also provided an optional bonus questionnaire (third section of the survey). 24 of 29 respondents answered bonus questions and received an additional $9 Amazon gift card. We provide the whole survey along with a sample story in the appendix.

We discuss our findings from the main survey in Sections 3.1 and 3.2 and from the bonus questions in Section 3.3. For such analysis, we deidentify the responses (by removing demographic attributes) and focus on understanding helpers' reactions to long stories and highlights.

3.1 Unhighlighted Long Stories are Demanding

For the first section of the survey, we selected ten stories (from 4933 relevant stories) longer than the mean length of the stories: 1881 characters. The mean length of these ten selected stories was 10 211 characters. Out of these ten stories, one story was randomly assigned to each respondent. Respondents were required to read the story and answer two questions: (i) how demanding was completely reading the story (Q1) and (ii) how demanding was it to construct a response (Q2).

On a Likert scale of one to four, the mean rating for Q1 was 3.13 (standard deviation of 0.95) and 3.10 (standard deviation of 0.90) for Q2. These high ratings indicate that helpers find unhighlighted stories demanding to read and reply to. However, in order to answer $\text{RQ}_{\text{demanding}}$, we need to compare these ratings with the demandingness of highlighted stories, as shown below.

3.2 Highlights are Helpful

In the second section of our survey, we showed respondents stories highlighted (based on our computational model's predictions about the incident, effect, and advice requested) and asked: (i) if the three highlighted parts provide important information about the survivor's situation and (ii) if the highlighted text assists helpers while reading and responding. We include all technical details pertaining to our dataset and model in Section 4.

For this human study, we selected ten stories different than the ones selected in the first section of the survey but of comparable lengths (average length was 10 135). We set up our survey in such a way that one of these ten highlighted stories was randomly assigned to each respondent. Respondents had to read the entire story containing highlights, write a supportive response addressing survivors' concerns, and answer the following survey questions:

Q3: Do the highlighted sentences represent important parts of the story?

Q4: (If answered yes to Q3) On a scale of one to four, how would you rate the importance of the highlighted text?

Q5: (If answered yes to Q3) What information did the highlighted text provide you?

Q6: How helpful was the highlighted text while reading and responding?

We did not want to overload each respondent in reading long texts as doing so may lead to poor quality of their responses. Hence, we gave each respondent two stories to read (one unhighlighted and one highlighted story). To have some overlap in the stories assigned to the respondents, we included 10 stories in our survey set.

Table 2 summarizes the results of our study. Almost all (28 of 29) respondents answered *yes* to question Q3, indicating the importance of the three parts we considered for highlighting. Moreover, Q4 received a high rating on the importance scale. In answering Q5, respondents backed their ratings with a qualitative text about the highlights' benefits. And, in Q6, they rated the predicted highlights as being highly helpful. Overall, Table 2 shows the effectiveness of our model in assisting helpers in providing support.

3.3 Demandingness is Reduced with Highlights

In the third section of the survey, we asked respondents the following questions:

Q7: On a scale of one to four, how demanding was it to read and understand the highlighted story?

Q8: On a scale of one to four, how demanding was it to construct a response to the highlighted story?

Q9: Are stories with highlights easier to read and understand than stories without highlights?

Q10: Is it easier to construct a response to stories with highlights than stories without highlights?

We analyzed if highlights affect the demandingness while reading and responding to long stories. To do so, we considered respondents' ratings for Q7 and Q8 and compared them with Q1 and Q2, respectively. For this comparison,

First section			
Q1	[1,4]	How demanding was completely reading the story?	Mean rating: 3.13, standard deviation: 0.95
Q2	[1,4]	How demanding was it to construct responses to the given story?	Mean rating: 3.10, standard deviation: 0.90
Second section			
Q3	Y/N	Do highlighted sentences represent important parts of the story?	28 of 29 responded Yes
Q4	[1, 4]	(If yes on Q3) How would you rate the importance of the highlighted text?	Mean rating: 3.34, standard deviation: 0.76
Q5	Text	(If yes on Q3) What information did the highlighted text provide you?	"Help me to focus on the keywords and better understand the content of this story" "Grabbed my attention" "It gives me a quick idea of what the point of the article is"
Q6	[1, 4]	How helpful was the highlighted text while reading and responding?	Mean rating: 3.37, standard deviation: 0.62
Third section			
Q7	[1, 4]	How demanding was it to read and understand the highlighted story?	Mean rating: 2.41, standard deviation: 1.01
Q8	[1, 4]	How demanding was it to construct a response to the highlighted story?	Mean rating: 2.41, standard deviation: 0.97
Q9	Y/N	Are stories with highlights easier to read and understand than stories without highlights?	23 of 24 responded Yes
Q10	Y/N	Is it easier to construct responses to stories with highlights than to stories without highlights?	23 of 24 responded Yes

Table 2: Key questions in our survey and summary of responses received. The complete survey is provided in the appendix.

we conducted a renowned statistical test namely, the Mann-Whitney U-test. In this test, we input all demandingness ratings with and without highlights. The statistical test shows that highlights significantly reduce demandingness in reading (U-statistic value $= 512$, p-value < 0.01) and responding (U-statistic value $= 210$, p-value < 0.02) to long stories. Additionally, it is important to note that after introducing highlights, there is about a 0.7 reduction (in mean value) on the four-point demandingness scale, again showing the importance of highlights and the benefit of our approach.

According to Q9 and Q10 responses, we found that almost all helpers (23 of 24) prefer highlighted stories as they are easier to read, understand, and respond to, than nonhighlighted stories.

4 MeThree Dataset and Classifier

We now describe the necessary technical details, starting from how we collected violence stories to how we applied active learning for highlighting. The active learning process developed hand-in-hand a labeled dataset (called MeThree) and a model to automatically highlight three parts of the story.

4.1 Collecting Sexual Violence Stories

We collected a total of 9140 violence stories from three subreddits: r/meToo, r/sexualassault, and r/SexualHarassment, for the period 2016-01-01 to 2021-07-18, using the Pushshift API (Pushshift 2023). This set of 9140 stories was the entire set of stories in that time range in those subreddits.

Some violence stories don't share survivors' experiences but instead, share news articles, seek opinions about allegations against celebrities, or promote other platforms. Such stories are irrelevant to our study. Similar to a previous study (Hassan et al. 2020), we applied the following heuristics to focus on stories containing survivors' personal experiences:

- First-person pronouns: Many survivors use first-person pronouns in the title of their story. For example, "**I** started to do something about **my** past assault, but instead of feeling better, it actually gets worse" and "**My** mom's boyfriend tried to get **me** to do things to him." Thus, we checked the presence of first-person pronouns: *i*, *me*, *my*, and *mine* in the title to find relevant violence stories.
- Advice-related keywords: We observed that survivors also use advice-related keywords in the title. For example, "Need **advice**, or support" and "pls someone read this and **help** me figure out if i was assaulted or not." We used the keyword, *advice*, as seed and queried its synonyms from the Oxford dictionary. We obtained 25 synonyms and selected four of them based on their relevance to our problem. We term these the *advice keywords*: *help*, *suggestion*, *advice*, *guide*, and *counsel*. To select relevant stories, we checked the presence of these keywords in the title. For extracting synonyms, we considered corpora such as WordNet (Miller 1995) but did not find synonyms that were commonly used.
- Advice-related questions: We observed that many rele-

vant stories ask a question (related to harassment or assault) in the title without mentioning any of the advice keywords. Example questions include "Was this rape?" and "Is this sexual harassment?". Using Part-Of-Speech (POS) tagging, the titles that have an interrogation form and include *rape*, *harassment*, *assault*, and *abuse* as the object, were selected.

Stories with titles satisfying one of the above rules were chosen. We checked 50 randomly selected stories for relevancy, involving a total of 812 sentences and a total 72 130 characters. Of these 50 stories, 47 (94%) were relevant because they either sought support or advice related to their case. Among these 47 stories, we found one story written by the survivor's friend but which described the effects on the survivor and sought advice.

We computed the recall of this filtering by analyzing stories (same sample size) not selected through the filtering method. The recall comes out to be 61.8%, for an F1 score of 74.6%. Other studies (Hassan et al. 2020; Ghosh Chowdhury et al. 2019b; Khatua, Cambria, and Khatua 2018) would have missed out on relevant data too as their data collection is similarly based on rules, keywords, or hashtags. But the objective here is to select relevant stories without further pruning—hence high precision is needed. Achieving high precision (94% in our case) means we can build a dataset of important sentences without further pruning.

In total, we obtained 4933 relevant stories using the above heuristics. Our statistics about mean and maximum length, and the distribution of the three parts pertain to this set of stories. Out of these 4933 selected stories, 74.29% (3665) are from r/sexualassault, followed by r/meToo (17.23%; 850), and r/SexualHarassment (8.47%; 418).

4.2 Using Active Learning for Efficient Labeling

The preparation of the dataset and the development of a classifier happen iteratively. For classification, we adopt pool-based active learning, which is known for training robust models while reducing manual labeling effort (Settles 2012). We tackle the problem of highlighting elements at the sentence level. Doing so gives us the ability to highlight key portions of long posts on Reddit.

We curate METHREE, a dataset comprising 8947 labeled sentences that are sampled from 4933 stories as explained below, and train an XLNet model on METHREE. Each sentence can either belong to None or at least one of three categories: incident, effects, and advice (being) requested. Hence, we have a multilabel classification task.

In active learning, the four steps shown in Figure 1 are repeated multiple times (Settles 2012). Pool-based active learning starts with an initial dataset (denoted by L) including labeled sentences, most of which were selected based on keywords (Section 4.3). In Figure 1, first, a model (denoted by M) is trained on the set L. To do so, we compared the performance of multiple models and chose the best-performing one (XLNet) as model M (Section 4.4). Second, an unlabeled dataset U is labeled by the predictions from the trained model M. In our case, since most of the sentences in L contained certain keywords, to avoid bias, we selected U from sentences without those keywords. Third, from U, data points having a high chance of being misclassified are queried and manually labeled (Section 4.5). Fourth, U is added to L (Section 4.6). We repeated the active learning cycle five times to curate METHREE, the final dataset of 8947 labeled sentences, and our computational model.

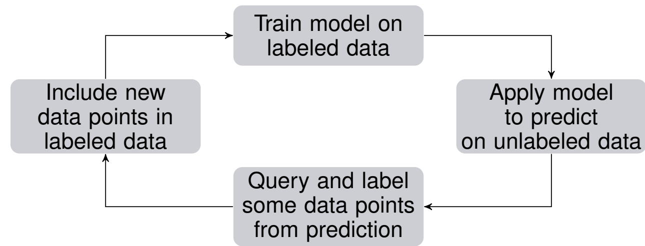


Figure 1: Active learning cycle involving four steps.

4.3 Initial Training Data for Active Learning

To form our initial training data (set L), we leverage 4933 relevant stories. First, for each category, we found candidate sentences from these stories. Second, we labeled a sample of candidate sentences along with other sentences.

Finding Candidate Sentences We split each of the 4933 relevant stories into 102 204 sentences, using Natural Language Toolkit (NLTK) library (NLTK 2023). However, a random sample of these sentences was inefficient in obtaining sentences that describe incidents, their effects, or advice being requested. Thus, we first found candidate sentences of each category using the following keywords:

- Incident: Hassan et al. (2020) created a list of 27 violence-related verbs (such as *molest*, *touch*, *rape*, *masturbate*). We expanded the list to 652 verbs by including synonyms of these verbs from the Oxford English Dictionary and then cut it down to the 539 most relevant verbs, of which only 313 were unique. We called this final set of 313 verbs *incident keywords*. We found 30 927 sentences having one or more incident keywords.
- Effects: We adopted four commonly expressed emotions—*anger*, *disgust*, *fear*, and *sadness*—from the NRC emotion lexicon (Mohammad and Turney 2013, 2010). We applied their associated lexicons and identified 37 271 emotional sentences. Next, we identified 14 synonyms of the word *feel* from the Oxford English Dictionary. Out of these, the following eight are relevant and form the *feelings keywords*: *feel*, *perceive*, *sense*, *experience*, *undergo*, *bear*, *endure*, and *suffer*. We found 8617 candidate sentences containing one or more of them.
- Advice requested: We observed that many questions in violence stories are advice seeking. Examples include "Was it actually just a mistake and should I forgive him?" and "Am I blowing it out of proportion?". Hence, we considered all questions as candidates for advice-seeking sentences. We found 6354 such candidates. Next, we leveraged the advice keywords from Section 4.1 to find an additional 2678 candidates.

To find synonyms, we used the Oxford English Dictionary as Corpora such as WordNet (Miller 1995) and PyDictionary (PyDictionary 2023) were not useful. For example, PyDictionary produced no synonyms representing feelings, and WordNet produced one word, *palpate*, which was uncommon to describe feelings.

Labeling Sentences Due to the presence of keywords (such as incident, feelings, and so on), candidate sentences are likely to be relevant to the three categories (incident, effects, and advice). Hence, for the set L, we took a random sample of 6900 sentences that match the above keywords. However, doing so would bias the training set (set L) toward these keywords. To mitigate that risk, at this step, we also included randomly selected 500 sentences that do not include any keywords. After discarding duplicates, we were left with a total of 5947 sentences.

Since a majority of these 5947 sentences contained our keywords, labeling them would produce a biased dataset. Note that this was the initial training data (set L). In subsequent steps of active learning, to mitigate bias, we kept including sentences without any keywords (set U), as described in Section 4.6. For these 5947 sentences, three of the authors of this paper were the annotators. Before labeling, they were aware of the uncomfortable and disturbing text present in these sentences. For each sentence, the annotators answered the following questions:

1. Does it describe sexual violence?
2. Does it describe the incident's effect on the survivor?
3. Does it indicate a request for advice?

The annotators read each sentence and answered the above questions as yes or no. Initially, two annotators labeled 200 sentences based on their understanding of the problem statement. Later, they discussed their disagreements and defined the final labeling instructions for all the annotators. Section 2 defines these three categories; detailed labeling instructions (along with examples) are in the appendix.

All 5947 sentences were divided among the three annotators (let's denote them by A_1, A_2, and A_3) such that each sentence was labeled by two of them. After labeling all the sentences, we obtained Cohen's kappa scores (Cohen 1960) of 0.772 (for incident), 0.774 (for effects), and 0.865 (for advice requested). These scores indicate that we achieved substantial agreement for two categories: incident and effects, and almost perfect agreement for the advice requested category. Table 3 also shows Cohen's kappa scores for each pair of annotators. Finally, the first author resolved all the disagreements. The labeled 5947 sentences form the initial training data (set L) for active learning.

For simplicity, annotators labeled independent sentences without seeing the surrounding sentences. Although Cohen's kappa scores are high, there is still a chance of mislabeling. Our study on randomly selected 370 sentences (sample size decided by Cochran's formula (Cochran 1977) using the standard values: 95% confidence, 0.05 margin of error, and 0.5 proportion of population) reveal that 98.19% of the labels (including all three categories) don't change even if we consider surrounding sentences during annotation. Hence, the risk of mislabeling is minimal in our case.

Annotators	Incident	Effects	Advice Requested
A_1, A_2	0.798	0.793	0.891
A_2, A_3	0.720	0.725	0.843
A_3, A_1	0.795	0.801	0.861
Total	0.772	0.774	0.865

Table 3: Cohen's kappa scores for each pair of annotators.

4.4 Initial Model to Highlight Sentences

A sentence acts as a concise and contiguous text conveying some information to the helper. Training a sentence-level classifier will yield concise highlights representing important sentences of the story. Hence, we treat our problem as sentence-level classification problem. Other highlighting solutions include predicting the boundary sentence (among consecutive sentences) where labels change and highlight the relevant ones. Such alternatives may capture more context than sentence-level classification. However, the varying nature of the consecutive sentences and the presence of multiple labels within a sentence pose challenges in predicting such boundary points in long stories. Hence, we preferred sentence-level classification over such alternatives.

Once set L is curated, the next step is to train a model M. This is a multilabel classification task in which each sentence is an input to the model and the output has three binary labels (one label for each category). Instances with all three labels as zero are predicted None. We trained and evaluated multiple methods on the 5947 labeled sentences (set L) as described below.

For each of these sentences, we computed embeddings such as Sentence-BERT (Reimers and Gurevych 2019), TF-IDF (Cahyani and Patasik 2021), Stanford's GloVe (Pennington, Socher, and Manning 2014), Word2Vec trained on the Google News (Mikolov et al. 2013), and Universal Sentence Encoder (USE) (Cer et al. 2018). For each embedding, we used the sentence vector as an input to a multilabel classifier. For GloVe and Word2Vec, we averaged word vectors to form the sentence vector. For classification, we tried Logistic Regression (LR) (Dreiseitl and Ohno-Machado 2002), Support Vector Machine (SVM) (Cervantes et al. 2020), and Random Forest (RF). We report the best method.

Besides embedding-based methods, we applied transformer-based approaches such as RoBERTa (Liu et al. 2019b) and XLNet (Yang et al. 2019). We fine-tuned RoBERTa and XLNet on set L by adding an output layer containing three units, one dedicated to each category. Both models minimized binary cross entropy over five epochs. The training batch size and tokenizer length were set to 32 and 256, respectively.

Table 5 in the appendix reports the mean F1, precision, and recall scores for the above approaches over ten-folds of set L, and compares these approaches with the keyword search (keywords used in Section 4.3). For the embeddings-based approaches, Table 5 reports the results only for their best-performing classifiers.

TF-IDF, GloVe, Word2Vec, Keyword search, and USE underperform as compared to other methods. Sentence-BERT with SVM achieves the highest macro precision (0.84). However, it shows lower macro recall (0.66) than RoBERTa (0.84) and XLNet (0.87). Overall, XlNet outperforms all other methods by achieving the highest macro F1 score (0.82). Thus, we choose XLNet as model M in our method.

4.5 Predicting and Querying from U

After model M is trained, we make predictions on the set U and label it. To mitigate the risk of a biased dataset, we chose the set U to be a random sample of 500 sentences (from relevant stories) not containing any keywords and applied the model M to it. We selected potentially misclassified sentences (in U) for manual labeling, as described below.

Uncertainty sampling (Culotta and McCallum 2005; Dagan and Engelson 1995) queries data points (sentences in our case) for labeling where the model is uncertain. However, uncertainty sampling methods (such as least confidence and entropy) did not work well in our case. This is because, in the first active learning cycle, model M (XLNet trained on 5947 sentences; Section 4.4) predicted low probabilities on sentences that lack the keywords (set U). We validated this by predicting on 100 such sentences, where the mean prediction probability was 0.08 (standard deviation of 0.23) for incident, 0.10 (standard deviation of 0.27) for effects, and 0.05 (standard deviation of 0.21) for advice requested. Due to these probabilities being low, uncertainty sampling methods could not discriminate between misclassified and other sentences.

To query the misclassified sentences from U for manual labeling, we found a threshold on the prediction probability. We discuss the query strategy in Appendix A.2. In the same strategy, we used a set U', another random sample of 500 sentences (not having any keywords) for validation and testing of identified thresholds.

4.6 Completing Active Learning Cycles

We completed the first active learning cycle (Figure 1) by adding labeled data U to set L. We iterated the cycle four more times, each time U having 500 sentences without keywords, to add 2500 labeled sentences to the 5947 initially labeled ones. As a result, the final L includes 8447 sentences. As stated in Section 4.5, we labeled U', an additional 500 sentences without keywords. After adding U', METHREE becomes a dataset of size 8947 labeled sentences.

In METHREE, there are 4331 (48.4%) sentences that belong to at least one category, and 4616 (51.6%) to None. Figure 2 shows the Venn diagram of the 4331 sentences over the three categories. As a result, the trained model performs multilabel classification.

Finally, we trained XLNet on METHREE to highlight sentences from long stories. Over ten cross validation, the model achieved a macro F1 score of 0.82 (0.78 for incident, 0.79 for effects, and 0.89 for advice requested), a macro recall of 0.86 (0.82 for incident, 0.83 for effects, and 0.92 for advice requested), and a macro precision of 0.78(0.74 for incident, 0.76 for effects, and 0.85 for advice requested).

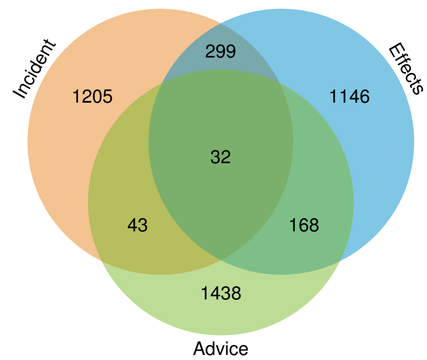


Figure 2: Distribution of sentences in METHREE across the three categories. A sentence in METHREE can belong to multiple categories (299 sentences for incident and effects; 168 sentences for advice and effects; 43 sentences for incident and advice; 32 sentences containing all three categories). Hence, training on METHREE for highlighting purpose is multilabel classification.

5 Potential Impact of Our Work

Our computational model was trained on sentences from three subreddits (r/meToo, r/sexualassault, and r/SexualHarassment). However, the realm of sexual violence stories extends to multiple other subreddits such as r/Molested, r/rape, r/rapecounselling, r/MenGetRapedToo, r/adultsurviviors, and r/afterthesilence, which offers a great opportunity to facilitate helpers' efforts in this community.

For real-world application, we envision our computational model to be fine-tuned for multiple violence-related subreddits. After fine-tuning, it can be applied on the stories posted every day (on subreddits) to highlight the important parts. Moreover, the model can be incrementally updated by incorporating new sentences posted on these subreddits. These highlights could be reviewed and approved by moderators, causing the stories to be shown to helpers with three types of highlighted text (incident, effects, and requested advice).

The generated highlights can potentially draw helpers' attention toward important parts, in turn assisting them in reading and responding to long stories (also shown through the survey). For example, focusing on the text describing mental state of the survivors (effects) would guide helpers to sensitively draft their responses. In addition, focusing on the requested advice section would guide them in making their responses customized to what survivors are looking for.

We also analyzed the output of violence stories from a summarization tool for comparison with our proposed strategy of highlighting text. For summarization, we applied the LED model, an encoder-decoder variant of Longformer model (transformer for long text), which has achieved state-of-the-art results on summarizing long documents (Beltagy, Peters, and Cohan 2020). Since we did not have ground truth summaries of 4933 relevant stories, we could not conduct quantitative analysis. Instead, we conducted qualitative analysis of these summaries.

Our qualitative analysis reveal that the LED model is able to provide the overview of the story yet misses some important parts. For example, we show below some excerpts from a summary that describe incident and effects but fail to include the advice requested. This is because most of the story revolves around the incident and effects, whereas advice is requested only in a few sentences. As a result, the summarizer considers incident and effects as the overview of the story. However, understanding the advice requested is crucial for providing effective help to a survivor (Andalibi et al. 2016; Rickwood et al. 2005; O'Neill 2018). We encountered a similar pattern of missing other key parts such as the incident.

Summary: *"... We both were sitting in the same room which was small and hence were almost touching. I was there for quite some time before he started making 'jokes'. He touched my hair and then face and said he was going to kiss me. This got me uncomfortable so I refused and asked to reverse what was happening. He tells me he was only joking and I believed him.... He started looking between my legs and pushing his legs up onto me. At that time, I simply ignored what was happening and made peace with it.... "*

On the other hand, our proposed approach is developed on highlighting three parts considered important in support-seeking literature. The following is an excerpt from the highlights (generated by our model) of the same story. In this case, the survivor's main purpose was to seek advice on validating their experience, but the summarizer could not identify advice questions at the end (highlighted below).

(We both were sitting in the same room which was small and hence were almost touching. ... He touched my hair and then face and said he was going to kiss me.) ... [This got me uncomfortable so I refused and asked to reverse what was happening.] ... (He started looking between my legs and pushing his legs up onto me.)... [I was very sad and weeping alone ... This was happening to me for the first time and wanted to share it with someone.]... <Am I overthinking this as an experience? ... Is it sexual harassment or nothing at all?>

That is, the LED longformer (used in our experiment) is able to provide good overviews of a story but misses out on parts that are crucial for support-seeking communities. Our approach includes those parts (based on the literature) and generates highlights aimed for assisting helpers.

6 Related Work

The literatures in psychology and healthcare focus on survivors' mental health (Gold et al. 2008; Fortier et al. 2009) and discuss how it can be improved (Rickwood et al. 2005; O'Neill 2018) through support seeking. Moreover, there has been extensive research in analyzing the stories seeking support and their reactions on online platforms (Manikonda et al. 2018; Deal et al. 2020; Field, Bhat, and Tsvetkov 2019; Reyes-Menendez, Saura, and Thomas 2020). However, these works don't primarily focus on highlighting tasks, unlike this work.

A few studies classify violence stories, whereas we focus on sentence-level classification for highlighting. Karlekar and Bansal (2018) collect 9892 violence stories from the SafeCity website (Safecity 2023) and classify them for one of the harassment types: (i) groping or touching, (ii) staring or ogling, and (iii) commenting. Other studies (Yan et al. 2019; Liu et al. 2019a) leverage the SafeCity dataset either for the same classification or for classifying other attributes such as the abuser's age (below 30 or older), the abuser's relationship with the survivor, and so on. Bauer et al. (2020) also use the SafeCity dataset and build a chatbot system to help survivors. The SafeCity dataset contains concise experiences (typically three or four sentences long) and is ill-suited to highlighting sentences, as we do.

Hassan et al. (2020) train a model on 520 761 tweets with the #MeToo hashtag to identify tweet-level attributes, such as the category of sexual violence reported, the survivor's identity (tweeter or not), and the survivor's gender. Ghosh Chowdhury et al. (2019b) label 5119 and classify tweets for types: (i) disclosure of a survivor's personal experience and (ii) nondisclosure. Studies by Khatua, Cambria, and Khatua (2018) and Ghosh Chowdhury et al. (2019a) also focus on similar classification tasks.

These studies identify relevant violence stories from a massive stream of social media text. The expectation is that a helper on these platforms can provide support to the survivor of the identified story. However, merely identifying relevant stories is not enough. To provide effective support, a helper should be able to read and understand those stories, which our sentence-level highlights facilitate.

Traditional text summarization works (Jadhav and Rajan 2018; Cheng and Lapata 2016; See, Liu, and Manning 2017; Li et al. 2011; Zhang, Li, and Gao 2012) are trained or evaluated on domains such as news but are not built for the sexual violence context. Our tool is intrinsically different because it highlights important parts instead of generating summaries.

Botelle et al. (2022) extract violence-related text from clinical healthcare records. They identify the presence and type of violence along with the status of the patient. However, their focus was not to extract other parts, such as the effects and advice requested. Moreover, the language used in medical records is typically different from informal language on social media. To the best of our knowledge, we are the first ones to analyze Reddit stories for highlighting purposes.

7 Discussion

We now discuss our conclusion, our work's limitations, and possible future directions.

7.1 Conclusion

Seeking and providing support is a crucial aspect of the online communities focused on sexual violence (Andalibi et al. 2016; Rickwood et al. 2005; O'Neill 2018). Our study addresses a challenge faced by helpers in these communities while reading and responding to long stories. The computational model developed is not meant to diminish the authenticity of survivors' voices or alter their narratives, nor is it intended to impose a singular narrative structure. Rather, the model is designed to spotlight three pivotal components of survivors' stories grounded in the literature.

The results of our work are encouraging, hinting at the potential benefits of highlighting important parts. The analysis of our sample demonstrated that stories with critical components highlighted exhibited a significant reduction in cognitive demand for helpers. In particular, our survey showed that on a four-point Likert scale, there is about a 0.7-point reduction in demandingness when stories were presented with highlights. This implies that the model succeeded in enhancing the readability and comprehensibility of survivor stories, potentially enabling helpers to respond more empathetically and efficiently. With the introduction of this model, we aim to complement and amplify the existing dynamics of support, fostering a more nuanced and empathetic dialogue regarding nonconsensual and other traumatic sexual experiences.

7.2 Limitations and Future Work

Our work suffers from some limitations, which motivate future improvements. One limitation is that our model is trained on the sentences collected from only three subreddits. We expect the nature of sentences in METHREE to be similar to sentences present on other violence-related subreddits. However, we would need to fine-tune the model before applying it to the other subreddits. Moreover, during the data collection process, the recall of the filtering step was low; raising it may help improve the effectiveness of our model.

Our work focuses on story structure but the content of the stories can be understood at greater depth than here. In a general sense, studies of moral postures taken by survivors or helpers (Xi and Singh 2024b) and blame assignment (Xi and Singh 2024a) are relevant to a deeper understanding of the social psychology at play. More specifically, narratives of trauma including sexual and domestic violence exhibit some important patterns (Saxena et al. 2025), which we might incorporate into improved versions of the present approach to help helpers and survivors.

For stories not receiving any human response, we can extend our work to generate a response computationally. Even if they are verified and edited by a human helper, generating such responses can enable helping a larger number of survivors.

It would help to conduct a larger-scale survey of helpers to assess multiple facets of violence stories. Through the survey, we can gauge factors (other than length) that influence improved engagement by helpers. For example, we could ask helpers if the type of abuse and the type of advice being requested influence the nature and extent of their engagement with the posted story.

Acknowledgments

Garg, Xi, and Yuan were with NC State University when this work was performed. We would like to thank the US Department of Defense (Science of Security Lablet grant) and the NSF (grant IIS-2116751) for supporting this research.

References

Andalibi, N.; Haimson, O.; Choudhury, M.; and Forte, A. 2016. Understanding Social Media Disclosures of Sexual Abuse Through the Lenses of Support Seeking and Anonymity. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 3906–3918. San Jose, California: Association for Computing Machinery.

Bauer, T.; Devrim, E.; Glazunov, M.; Jaramillo, W. L.; Mohan, B.; and Spanakis, G. 2020. #MeTooMaastricht: Building a Chatbot to Assist Survivors of Sexual Harassment. In *Machine Learning and Knowledge Discovery in Databases*, 503–521. Cham: Springer International Publishing.

Beltagy, I.; Peters, M. E.; and Cohan, A. 2020. Longformer: The Long-Document Transformer. *arXiv:2004.05150*.

Botelle, R.; Bhavsar, V.; Kadra-Scalzo, G.; Mascio, A.; Williams, M. V.; Roberts, A.; Velupillai, S.; and Stewart, R. 2022. Can Natural Language Processing Models Extract and Classify Instances of Interpersonal Violence in Mental Healthcare Electronic Records: An Applied Evaluative Study. *BMJ Open*, 12(2): e052911.

Cahyani, D. E.; and Patasik, I. 2021. Performance Comparison of TF-IDF and Word2Vec Models for Emotion Text Classification. *Bulletin of Electrical Engineering and Informatics*, 10(5): 2780–2788.

Cer, D.; Yang, Y.; Kong, S.; Hua, N.; Limtiaco, N.; John, R. S.; Constant, N.; Guajardo-Cespedes, M.; Yuan, S.; Tar, C.; Sung, Y.; Strope, B.; and Kurzweil, R. 2018. Universal Sentence Encoder. *CoRR*, abs/1803.11175: 1–7.

Cervantes, J.; Garcia-Lamont, F.; Rodriguez-Mazahua, L.; and Lopez, A. 2020. A Comprehensive Survey on Support Vector Machine Classification: Applications, Challenges and Trends. *Neurocomputing*, 408: 189–215.

Cheng, J.; and Lapata, M. 2016. Neural Summarization by Extracting Sentences and Words. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 484–494. Berlin, Germany: Association for Computational Linguistics.

Cochran, W. G. 1977. *Sampling Techniques*. John Wiley & Sons Inc.

Cohen, J. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1): 37–46.

Culotta, A.; and McCallum, A. 2005. Reducing Labeling Effort for Structured Prediction Tasks. In *American Association for Artificial Intelligence*, volume 5, 746–751.

Dagan, I.; and Engelson, S. P. 1995. *Committee-Based Sampling for Training Probabilistic Classifiers*, 150–157. Elsevier.

Deal, B.-E.; Martinez, L. S.; Spitzberg, B. H.; and Tsou, M.-H. 2020. "I Definitely Did Not Report It When I Was Raped. . . # WeBelieveChristine# MeToos": A Content Analysis of Disclosures of Sexual Assault on Twitter. *Social Media+ Society*, 6(4): 2056305120974610.

Dreiseitl, S.; and Ohno-Machado, L. 2002. Logistic Regression and Artificial Neural Network Classification Models: a Methodology Review. *Journal of Biomedical Informatics*, 35(5): 352–359.

EEOC. 2023. Harassment. United States Equal Employment Opportunity Commission. Accessed 2024-06-19.

Field, A.; Bhat, G.; and Tsvetkov, Y. 2019. Contextual Affective Analysis: A Case Study of People Portrayals in Online #Metoo Stories. In *Proceedings of the 13th International AAAI Conference on Web and Social Media*, volume 13, 158–169.

Field-Springer, K.; Draut, H.; Babrow, F.; and Sandman, M. 2022. (Re)claiming Stories in the #MeToo Movement: Righting Epistemic Wrongs of Physical, Mental, and Emotional Harms of Sexual Violence. *Health Communication*, 37(8): 982–991.

Fortier, M. A.; Dilillo, D.; Messman-Moore, T. L.; Peugh, J.; DeNardi, K. A.; and Gaffey, K. J. 2009. Severity of Child Sexual Abuse and Revictimization: The Mediating Role of Coping and Trauma Symptoms. *Psychology of Women Quarterly*, 33(3): 308–320.

Gartner, R. E.; and Sterzing, P. R. 2016. Gender Microaggressions as a Gateway to Sexual Harassment and Sexual Assault: Expanding the Conceptualization of Youth Sexual Violence. *Affilia*, 31(4): 491–503.

Ghosh Chowdhury, A.; Sawhney, R.; Mathur, P.; Mahata, D.; and Ratn Shah, R. 2019a. Speak up, Fight Back! Detection of Social Media Disclosures of Sexual Harassment. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, 136–146. Minneapolis, Minnesota: Association for Computational Linguistics.

Ghosh Chowdhury, A.; Sawhney, R.; Shah, R. R.; and Mahata, D. 2019b. #YouToo? Detection of Personal Recollections of Sexual Harassment on Social Media. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2527–2537. Florence, Italy: Association for Computational Linguistics.

Gold, S. N.; Ketchman, S. A.; Zucker, I.; Cott, M. A.; and Sellers, A. H. 2008. Relationship Between Dissociative and Medically Unexplained Symptoms in Men and Women Reporting Childhood Sexual Abuse. *Journal of Family Violence*, 23: 569–575.

Hassan, N.; Poudel, A.; Hale, J.; Hubacek, C.; Huq, K. T.; Karmaker Santu, S. K.; and Ahmed, S. I. 2020. Towards Automated Sexual Violence Report Tracking. In *Proceedings of the 14th International AAAI Conference on Web and Social Media*, 250–259. Georgia, USA.

Jadhav, A.; and Rajan, V. 2018. Extractive Summarization with SWAP-NET: Sentences and Words from Alternating Pointer Networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 142–151. Melbourne, Australia: Association for Computational Linguistics.

Karlekar, S.; and Bansal, M. 2018. SafeCity: Understanding Diverse Forms of Sexual Harassment Personal Stories. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2805–2811. Brussels, Belgium: Association for Computational Linguistics.

Khatua, A.; Cambria, E.; and Khatua, A. 2018. Sounds of Silence Breakers: Exploring Sexual Violence on Twitter. In *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 397–400. Barcelona: IEEE Press.

Li, P.; Wang, Y.; Gao, W.; and Jiang, J. 2011. Generating Aspect-Oriented Multi-Document Summarization with Event-Aspect Model. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 1137–1146. Edinburgh, Scotland, UK.: Association for Computational Linguistics.

Liu, Y.; Li, Q.; Liu, X.; Zhang, Q.; and Si, L. 2019a. Sexual Harassment Story Classification and Key Information Identification. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2385–2388. New York, NY, USA: Association for Computing Machinery.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019b. RoBERTa: A Robustly Optimized Bert Pretraining Approach. *arXiv preprint arXiv:1907.11692*.

Manikonda, L.; Beigi, G.; Kambhampati, S.; and Liu, H. 2018. #Metoo through the Lens of Social Media. In *Proceedings of the 11th International Conference on Social, Cultural, and Behavioral Modeling*, 104–110. Washington, USA: Springer Verlag.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.; and Dean, J. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS, 3111–3119. Lake Tahoe, Nevada: Neural Information Processing Systems Foundation.

Miller, G. A. 1995. WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11): 39–41.

Mohammad, S.; and Turney, P. 2010. Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. In *Proceedings of the NAACL HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 26–34. Los Angeles, California: Association for Computational Linguistics.

Mohammad, S. M.; and Turney, P. D. 2013. Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence*, 29(3): 436–465.

NLTK. 2023. Natural Language Toolkit. NLTK Library. Accessed 2024-06-19.

NSVR. 2023. Harassment Statistics. National Sexual Violence Resource Centre. Accessed 2024-06-19.

O'Neill, T. 2018. 'Today I Speak': Exploring How Victim-Survivors Use Reddit. *International Journal for Crime, Justice and Social Democracy*, 7(1): 44–59.

Pennington, J.; Socher, R.; and Manning, C. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. Doha, Qatar: Association for Computational Linguistics.

Pushshift. 2023. Pushshift API. Reddit's Pushshift API. Accessed 2024-02-19.

PyDictionary. 2023. Dictionary in Python. Accessed 2024-01-10.

RAINN. 2023. Assault. Rape, Abuse & Incest National Network. Accessed 2023-07-19.

Reddit. 2023a. Subreddit MeToo. MeToo Subreddit. Accessed 2023-09-20.

Reddit. 2023b. Subreddit Sexual Assault. Sexual Assault Subreddit. Accessed 2023-04-20.

Reddit. 2023c. Subreddit Sexual Harassment. Sexual Harassment Subreddit. Accessed 2023-04-20.

Reimers, N.; and Gurevych, I. 2019. Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3982–3992. Hong Kong, China: Association for Computational Linguistics.

Reyes-Menendez, A.; Saura, J. R.; and Thomas, S. B. 2020. Exploring Key Indicators of Social Identity in the #MeToo Era: Using Discourse Analysis in UGC. *International Journal of Information Management*, 54: 102129.

Rickwood, D.; Deane, F. P.; Wilson, C. J.; and Ciarrochi, J. 2005. Young People's Help-Seeking for Mental Health Problems. *Australian e-Journal for the Advancement of Mental Health*, 4(3): 218–251.

Safecity. 2023. Safecity Website. Safecity Website. Accessed 2024-02-14.

Saxena, M.; Garg, V.; Ray, B.; Mishra, A.; and Singh, M. P. 2025. Understanding Narratives of Trauma in Social Media. In *Proceedings of the 17th ACM Web Science Conference (WebSci)*, 1–10. New Brunswick, New Jersey: ACM.

See, A.; Liu, P. J.; and Manning, C. D. 2017. Get To The Point: Summarization with Pointer-Generator Networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1073–1083. Vancouver, Canada: Association for Computational Linguistics.

Settles, B. 2012. *Active Learning*. Number 6 in Synthesis Lectures on Artificial Intelligence and Machine Learning. Cham, Switzerland: Morgan & Claypool.

Xi, R.; and Singh, M. P. 2024a. The Blame Game: Understanding Blame Assignment in Social Media. *IEEE Transactions on Computational Social Systems (TCSS)*, 11(2): 2267–2276.

Xi, R.; and Singh, M. P. 2024b. Morality in the Mundane: Categorizing Moral Reasoning in Real-life Social Situations. In *Proceedings of the 18th International AAAI Conference on Web and Social Media (ICWSM)*, 1648–1660. Buffalo, New York: AAAI Press.

Yan, P.; Li, L.; Chen, W.; and Zeng, D. 2019. Quantum-inspired Density Matrix Encoder for Sexual Harassment Personal Stories Classification. In *IEEE International Conference on Intelligence and Security Informatics, ISI*, 218–220. Shenzhen, China: Institute of Electrical and Electronics Engineers Inc.

Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R. R.; and Le, Q. V. 2019. XLNet: Generalized Autoregressive Pretraining for Language Understanding. *Advances in Neural Information Processing Systems*, 32.

Youden, W. J. 1950. Index for Rating Diagnostic Tests. *Cancer*, 3(1): 32–35.

Zhang, R.; Li, W.; and Gao, D. 2012. Generating Coherent Summaries with Textual Aspects. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, 1727–1733. Toronto, Ontario, Canada: AAAI Press.

Ethics Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, by conducting the survey
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes
   (e) Did you describe the limitations of your work? Yes, in Section 7
   (f) Did you discuss any potential negative societal impacts of your work? Yes, in Section 7.2
   (g) Did you discuss any potential misuse of your work? NA to the best of our knowledge
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, in Section 7.2
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? NA
   (b) Have you provided justifications for all theoretical results? NA
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? NA
   (e) Did you address potential biases or limitations in your theoretical framework? NA
   (f) Have you related your theoretical results to the existing literature in social science? NA
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? NA
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes, submitted as a supplementary file
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes, in Section 4 and Appendix
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes, through the survey conducted
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? NA
   (b) Did you mention the license of the assets? NA
   (c) Did you include any new assets in the supplemental material or as a URL? Yes
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes, in Section 3 for survey
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes, in Section 7.2.
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? Yes, in Section 7.2
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset ? Yes
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? Yes, in Appendix A.3
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? Yes, in Section 3
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? Yes, in Section 3
   (d) Did you discuss how data is stored, shared, and deidentified? Yes, in Section 3

Ethical Considerations

We describe our critical processes below and their associated risks from an ethical standpoint.

1. **Consent:** Our data was collected from public Reddit stories. We did not obtain consent of the survivors writing such stories. Also, some survivors may wish not to be contacted (Ghosh Chowdhury et al. 2019b).
2. **Anonymity:** We did not save survivors' personal information such as usernames. We paraphrased the stories and sentences presented in this paper as examples to remove potentially identifying content, such as the survivor's age, job title, and location.
3. **Labeling disturbing text:** Text taken from stories about violence can be disturbing. Therefore, we didn't hire crowd workers or volunteers for labeling; instead, three authors of this paper carried out the labeling.
4. **Potential misinterpretation:** We were extremely aware of the sensitivity of this research before labeling sentences. However, we may have misinterpreted some sexual violence experiences. That's why we don't claim that our labeling is fully accurate.
5. **Reproducibility and research:** We are not posting METHREE openly to avoid the risk of drawing attention to personal stories. We will share its anonymized version (along with metadata) upon request by faculty members or senior researchers (but not unsupervised students) at accredited institutions.

Our highlighting approach presents survivors' full stories to helpers, unlike summarization approaches that show only concise summaries—such summarization may appear to invalidate survivors' experiences.

A Appendix

A.1 Labeling Instructions

The final labeling instructions are described below:

1. Incident: Any unwelcome sexual advances, sexual behavior, requests for sexual favors, verbal or physical acts of sexual nature, offensive jokes, or remarks that were either sexual or based on someone's gender were labeled as incidents (Gartner and Sterzing 2016; EEOC 2023; RAINN 2023).
2. Effects on the survivor: We considered survivors' feelings and emotions that arose during or after the incident. Examples range from feeling uncomfortable (due to the abuser's actions) to being afraid (emotion: fear) of reporting incident.
3. Advice requested: We considered sentences in which survivors asked for suggestions on topics related to incident, e.g., whether to report the incident, where to get therapy from, and how to face the abuser again.

Table 4 shows examples of each category. For confidentiality, we have masked details that might be used to identify survivors or abusers. The first example describes inappropriate physical behavior and is considered an incident. The second example describes that the survivor is sexually exploited (incident) and suffers from depression and anxiety (effects). In the third example, the survivor expresses fear (by mentioning "freak out") and seeks advice about dealing with it. In the last example, the survivor seeks advice relating to the legal process.

A.2 Query Strategy

To find the threshold to query sentences, we used a set U', another random sample of 500 sentences not having any keywords. On U', we plotted the Receiver Operating Characteristic (ROC) curve and computed Youden's J-statistic (Youden 1950), as described below. This plot was based on the first active learning cycle and the model M referred to below is XLNet trained on the 5947 sentences described in Section 4.4.

1. The first author labeled the set U'.

   One annotator was adequate because of the high inter-rater agreement that we achieved on the earlier sample of 5947 sentences
2. The model M made predictions on the set U'.
3. We split U' into 400 sentences (set V) and 100 sentences (set T).
4. We used V to fine-tune the threshold above which we would select sentences for manual labeling. Out of the above 400 sentences, model M misclassified 28 sentences for the incident category, 38 for effects, and 9 for advice requested. Since we wish to retrieve these sentences, we consider misclassification as positive and plot its true positive rate on the Y-axis of the ROC curve for each category, as Figures 3a–3c show. From each ROC curve, we found the best threshold by maximizing Youden's J-statistic (Youden 1950), which equals the height above the chance line.

   For incidents, effects, and advice, we identified thresholds of $0.038\,177$, $0.008\,476$, and $0.007\,874$, respectively. We selected sentences in the classes if they exceeded the respective threshold.
5. For each category, to ensure that we did not miss out on the misclassified sentences, we additionally selected 30 sentences below the threshold for every 100 predictions. For example, the model M predicted on the set V that has $4 \times 100 = 400$ sentences. Therefore, we selected 4×30 = 120 sentences for each category below the threshold.

   In brief, for each category, query (1) sentences with prediction probability above or equal to the threshold, and (2) 30 *sentences below the threshold for every* 100 *sentences.*

   Through this query strategy, we could identify most of the misclassified sentences in V: 89.28% for the incident, 92.10% for effects, and 100% for advice requested.
6. We used the set T to test our query method. In T, M misclassified 10 sentences in the incident, four in effects, and three in advice. Our thresholds identified nine (90%) misclassified incident sentences and all misclassified sentences (100%) for effects and advice.

Sentence	I	E	A	Rationale
... he slid his hand up my leg and into my shorts.	✓			Inappropriate physical behavior
... I was sexually used by <abuser> on many occasions ... I am in a constant battle with major depression, crippling real event OCD (I ruminate for 16 hours/day) & debilitating anxiety.	✓	✓		Sexual abuse leading to depression and anxiety
... I'm freaking out and have no one to talk to because no one knows about him or what happened ... What do I do?		✓	✓	Freaking out and asking about future action
Does anyone know how a legal advocate works and what you experienced with them?			✓	Asking for legal advice

Table 4: Relevant examples according to the labeling instructions. Columns I, E, and A reflect the incident, effects, and advice requested, respectively.

We used the above method in each iteration to retrieve potential misclassified sentences from U and manually labeled retrieved sentences. For manual labeling, each of the three annotators (authors of this paper) labeled the sentences retrieved for a category.

A.3 Survey Details

In our survey, respondents filled the consent form followed by answering the following questions:

- **Demographics**
- How many responses do you think you have written in total on sexual violence-related subreddits? Please take your best guess.
  - — No responses
  - — Between 1 and 5
  - — Between 6 and 10
  - — Between 11 and 30
  - — More than 30
- What is your age? (optional to answer)
- What is your gender? (optional to answer)
- **First section**
- Please read the following real story from a sexual violence-related subreddit, and write a helpful or supportive response for the survivor.

  <One sample story was shown>

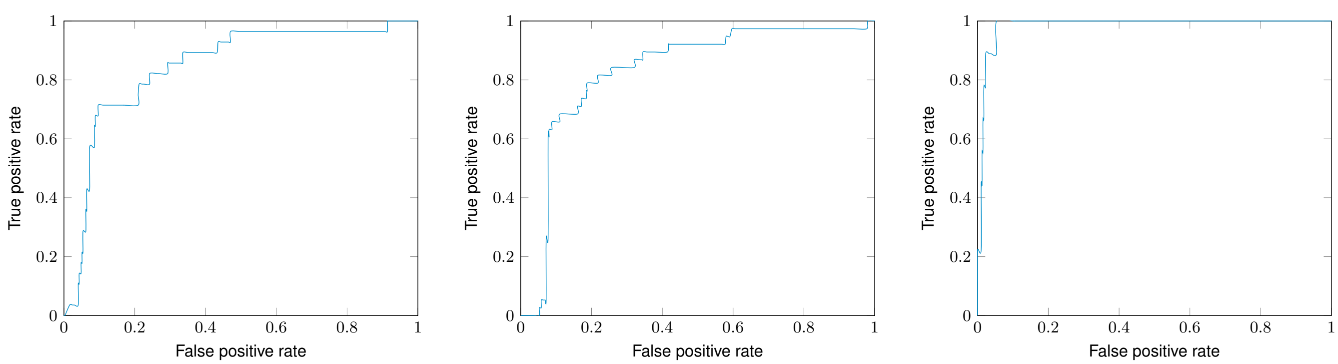
Sample story presented in survey.

At <age>, I was studying in <school-name> when I was also dating my first boyfriend. He was <age> at that time. We never made out and had no experience even with kissing or other things. None in my group used to like him. He was very bad at his studies and failed all subjects in school (he even repeated some years) but was good at sports. he visited me quite a few times when I was in boarding school. I could see some red flags in him but I did not bother. This was because he was too affectionate and touchy-feely, he would put his hand above mine and I would move away, he held my hand without consent so I pulled out mine, and he asked if he could hug me, then hugged me without me answering anything. I felt uncomfortable even if the incident was not sexual. I used to ignore all this but later realized I should have kept my boundaries. Once I was taking the train from a group outing to <place> for a family occasion. He used to live near the place and suddenly appeared to see me after the function. Then he gave an excuse that he wanted something to buy nearby, hence he boarded the train with me. Despite me refusing multiple times, he was waiting for me while I was meeting my dad at the station. I was seriously feeling uncomfortable but he wouldn't leave until my dad showed up and he said hi to my dad, my dad even questioned me who he was to which I replied that he is a friend. These incidents are before we started dating. After a few months into the relationship, when we used to spend time and eventually took a nap, he would touch me inappropriately on my private parts with my clothes on. It was not in an aggressive way, but he was wrong in his assumption that I was sleeping. I didn't have courage to confront at that moment as it was shocking and confusing experience, but later on I did confront him through chats to which he denied all accusations (and me putting lot of pressure, he finally felt apologetic and convey that he was sorry for being opportunist at that moment). By nature, he was also clingy on and possessive, and whenever I had to to go somewhere (such as to and from my workplace) he would check on me by coming along. Initially, I considered these incidents as his sweet gestures, such that he was investing all his time and effort to keep seeing me and worries for my well-being, but later on I realized he was going overboard and didn't like being around him all the time. After a few months into our relationship, I was almost <minor's age> (and him almost <age>), he used to bring up the topic of having sex and kept persuading me for the same. I refused his sexual offers a few times so he didn't do much physically. But one day I finally gave up on his insistence and we had sex on my bed in one of my rooms. He was quite considerate and was very cautious to make sure I wouldn't get pregnant (used contraceptives) but didn't know anything about foreplay so it was not there. And gosh did it hurt! After that, we started having sex quite often, almost every time when he came to visit me. I turned 17 by then. Slowly my pain was reducing and each time he used contraceptives but there was still little to no foreplay. Once he suggested me to go to a hotel with him for having sex and I did not refuse, which shows how dumb I am. We ended up going to a hotel multiple times, all during the initial parts of the day except once when it was a special occasion and we made out at night at that time. To my surprise, we did sex multiple times in the same night. After the first few times of having sex, he stopped asking for consent, I never convey my yes to him for the following "n" times, he just assumed it and inserted his penis every time without asking. I would admit that I did not struggle or yell, but each of the times it was physically hurting and he would not stop until he finishes. I never confronted even once with him. At one point during my final years of study, I was highly anxious as I might have gone pregnant because contraceptives also have a failure rate. Moreover, once when he pulled his penis out after he finished and he was deflating so the contraceptive didn't come out with his penis, hence he used his hand to pull it out. Since I am an overthinker and worry a lot, so I started becoming anxious. I took 2 pregnancy tests and both were negative. My private part was also bleeding a bit so I decided to consult the doctor. The doctor said that this case is of implantation bleeding which generally happens after a few days of period. Due to God's grace, I was not pregnant at that time, hence I also received my next period the next month. The next year of our relationship went into a lot of arguments, we argued in the worst possible ways and I initiated to break this tie with him. I met him at <place> and brought all the gifts and other things that he gave me but he accepted none of them. He was extremely angry at me and I was afraid if he would yell at me in public, so I took him to the roof of that place. he then shouted at me and threw all the stuff around. He also bought some liquid food items for eating together but he took one cup and poured it all over his clothes and body. After this incident, I was extremely shocked, taken aback, and afraid at the same time. As a result, I just froze. By the time he went to one of the restrooms to clean his clothes and found me huddled on the ground after coming back. After seeing me this way, he hugged me and apologized for his behavior. Prior to this incident, he would threaten to hurt himself during any argument. Once, he threatened to jump off the roof (ran towards the edge and then stopped), took the scissors and a nail clipper and acted to hurt his fingers and hands, and even took a knife and pretended to hurt himself when I visited him at his home. As a result, I was worried and concerned about him and took all potential-hurting equipments away from him, but it was that concern he wanted to see in me, to confirm my care towards him. By the way, he also used to show care for me and was always there for me as my boyfriend, especially in investing some time just to see that I am okay. He also bought me some food items such as snacks and drinks because he knew that I like eating all of them, he would always ask for my heavy bag whenever I am carrying it, always opened doors for me, and was always with me whenever I took public transport anywhere. On the day of <occasion>, I told him not to come because my family would be present and there were only two people allowed to accompany the candidate. When I went back to the dorms, he was already there waiting for me. Then he wouldn't allow me to have lunch in the cafeteria because he himself was hungry and brought some food for both of us to eat in my room. He was already angry with me because I did not allow him to attend the <occasion>. I was still wearing my outfit for the occasion, and suddenly I see him lifting my outfit and inserting his penis inside me. I was not in the mood and refused to have sex at that time. However, he did not listen and kept going until he ejaculated. At that moment, I never thought that he raped me in my own room and that too in an outfit. He even said I was ungrateful because I did not thank him enough for the food that he ordered for both of us. I generally say thanks, hence he was expecting it this time. I told him that he was the one who restricted me from eating at the cafeteria. After I graduated, I went to a different country for my higher education. Even though he was not with me, we were still in touch with each other through online chats. During winter break, I flew back to visit my home and met family and friends. He came to pick me up at the airport and helped with my luggage which was nice of him. But when we arrived at my home, he wanted to come in just to see my family and interact with them. I was very confident that I do not want him to know which floor I lived so I asked him to go up to my front door where I literally shouted at him and asked him to leave. During the break, I visited his home once. He asked me to come to his room when his family including his parents were not home. I was happy and surprised that he had bought cloth for me that I liked, so I even thanked him for that. Then he instructed me to lie down on his bed, I thought we were going to cuddle so I did lie down. He climbed into bed after me and suddenly started taking my pants off. I was very perplexed by his reaction and put my hand out to stop him but he did not resist and continued taking my pants off. After a point, he disappeared from the bed for a while and I wondered where he ran to. Still oblivious to the danger I was in. When he came back he had contraceptives with him, forced me down on the bed, and inserted his penis suddenly. I shouted at him and told him it hurts. He said it's hurting because we have not done it in the past few months just because I went abroad for my studies. He then continued until he got the satisfaction that he required. After a few months, I finally broke up our relationship because I found he cheated on me since I flew back abroad. At first, he refused all allegations but then got desperate because I was breaking up with him and admitted that he cheated on me quite a few times. Still, I was very confident that I would never go back to him again and even told him that. I blocked him from all of my social media accounts. He sent me 3 parcels in which he was saying sorry with small gifts inside and I briefly unblocked him to tell him to stop sending me things or bothering me and my social circle before blocking him again. When I flew back home during another vacation, I bumped into him twice in public and felt extremely uncomfortable. Throughout those times he had been desperately trying to make contact through some new and fake social media accounts and again trying to connect with me and my friends. He texted some of my friends who would get back to me regarding this. One time he even went to my friend's place, called her and pushed her to meet him. She was scared and went down to meet him, where she gave him some ideas of what to gift me to apologize again. He even reached out to my mom pushing to meet her for food and wanting to buy me a flight to fly back due to covid era. I was absolutely stressed and sad that he was reaching out to my family and multiple friends and I couldn't do anything about it except guide them on how to handle him. There were many more wrong things he did, these were just the main things I felt that forced me to have an opinion against him. After a whole year, he was still reaching out to talk to me, demanding why I blocked him and trying to prove he did nothing and was not even a cheater. He turned the tables on me saying that he said that he cheated just to see my reaction, to which he found me going out with a guy friend for a drink once. Whenever I think of all this, I feel like I was so screwed up during all those years I spent with him. I never ate healthy food, hardly studied anything, and was constantly under pressure and anxiety. Now, I feel I am safe and in a much better place in all aspects: physically, emotionally, mentally, spiritually, and psychologically. Recently, I am dating a guy in school and he is also the only person who knows everything about my past assault. I know that if I told my parents about my ex, they would be highly stressed and restrict me even for going out. I don't get support from my friend. They think that I was still with my ex even when I did not like him. But they do not know the whole matter. I m very fortunate that I found someone now who understands me, connects well, guide and love me. Thanks to god!

- How demanding was completely reading the story?
  — Not much demanding
  — A little bit demanding
  — Quite a bit demanding
  — Significant demanding
- How demanding was it to construct responses to the given story?
  — Not much demanding
  — A little bit demanding
  — Quite a bit demanding
  — Significant demanding
- **Second section**
- Now, you will do the same thing you just did. Please read the following story and then write a helpful or supportive response for the survivor.
- <One violence story with highlights is shown>
- The post you just read contained highlighted sentences. Do these sentences represent some important parts of the story?
  — Yes
  — No
- (If yes to the last question) How would you rate the importance of the highlighted text?
  — Not much important
  — A little bit important
  — Quite a bit important
  — Significant important
- (If yes to the second last question) What information did the highlighted text provide you?
- How helpful was the highlighted text in reading and responding?
  — Not much helpful
  — A little bit helpful
  — Quite a bit helpful
  — Significant helpful
- **Third section**
- How demanding was it to read and understand the highlighted story?
  — Not much demanding
  — A little bit demanding
  — Quite a bit demanding
  — Significant demanding
- How demanding was it to construct a response to the highlighted story?
  — Not much demanding
  — A little bit demanding
  — Quite a bit demanding
  — Significant demanding
- Are stories with highlights easier to read and understand than stories without highlights?
  — Yes
  — No
- Is it easier to respond to stories with highlights than to stories without highlights?
  — Yes
  — No

Category	Metric	TF-IDF+ LR	GloVe + RF	Word2Vec + RF	Keyword search	USE + SVM	Sentence-BERT + SVM	RoBERTa	XLNet
Incident	F1 Score	0.64	0.43	0.50	0.47	0.69	0.70	**0.77**	**0.77**
	Recall	0.56	0.31	0.33	0.79	0.63	0.61	0.80	**0.82**
	Precision	0.75	0.71	0.73	0.34	0.77	**0.83**	0.74	0.74
Effects on the survivor	F1 Score	0.66	0.38	0.39	0.49	0.69	0.70	0.78	**0.80**
	Recall	0.60	0.25	0.26	0.82	0.65	0.61	0.80	**0.86**
	Precision	0.73	0.76	0.79	0.35	0.75	**0.84**	0.77	0.74
Advice requested	F1 Score	0.74	0.63	0.64	0.67	0.74	0.80	**0.89**	**0.89**
	Recall	0.70	0.58	0.59	0.93	0.71	0.75	0.91	**0.94**
	Precision	0.78	0.73	0.73	0.53	0.77	0.86	**0.87**	0.84
Macro	F1 Score	0.68	0.48	0.51	0.54	0.71	0.73	0.81	**0.82**
	Recall	0.62	0.38	0.39	0.85	0.66	0.66	0.84	**0.87**
	Precision	0.75	0.73	0.75	0.41	0.76	**0.84**	0.80	0.77

Table 5: Comparing performance of multiple trained models. Bold indicates the highest score for a metric.



(a) Incident category. (b) Effects category. (c) Advice requested category.

Figure 3: ROC curves showing true positive and false positive rates, while considering misclassified sentences under positive class.