BU-NEmo: an Affective Dataset of Gun Violence News

Carley Reardon, Sejin Paik, Ge Gao, Meet Parekh,
Yanling Zhao, Lei Guo, Margrit Betke, Derry Wijaya

Boston University

{reardonc, sejin, ggao02, meet09, lingzhao, guolei, betke, wijaya}@bu.edu

Abstract

Given our society's increased exposure to multimedia formats on social media platforms, efforts to understand how digital content impacts people's emotions are burgeoning. As such, we introduce a U.S. gun violence news dataset that contains news headline and image pairings from 840 news articles with 15K high-quality, crowdsourced annotations on emotional responses to the news pairings. We created three experimental conditions for the annotation process: two with a single modality (headline or image only), and one multimodal (headline and image together). In contrast to prior works on affectively-annotated data, our dataset includes annotations on the dominant emotion experienced with the content, the intensity of the selected emotion and an open-ended, written component. By collecting annotations on different modalities of the same news content pairings, we explore the relationship between image and text influence on human emotional response. We offer initial analysis on our dataset, showing the nuanced affective differences that appear due to modality and individual factors such as political leaning and media consumption habits. Our dataset is made publicly available to facilitate future research in affective computing.

Keywords: affective computing, multimodal news data, crowdsourcing, language resources

1. Introduction

Through the rise in popularity of online media consumption in the last decade (2011-2021), digital content types are more varied than ever with increased multimedia experiences (Westcott et al., 2021; Shearer, 2021). In particular, a wide-range of research exists on the emotional impact of online, multimodal content. Examples of such work include measuring the emotional impact of social media post shares (Gupta and Yang, 2019), analyzing the influence of partisanship in affective political news (Hasell, 2020), and uncovering high-frequency emotions in fake news (Zhang et al., 2021b).

Within machine learning, the task of affective emotion recognition aims to predict the emotional response that a given piece of media will elicit. While much work exists to benchmark the basic performance of this task, there is a lack of news media data for this task, with most notable datasets collecting abstract emotional responses through art, such as Achlioptas et al. (2021). Some use social and/or news media, but they neglect to fully annotate the data with detailed emotional responses, as in You et al. (2016), Yang et al. (2017), and Guo et al. (2021).

This study focuses on gun-violence news as it continues to significantly impact the lives of everyday people, with an average of 108 people in the U.S. dying daily from gun-related deaths (Lin, 2021) and federal data showing the rise of gun sales during the COVID-19 pandemic (Schaeffer, 2021). Despite this, gun violence and gun control continue to be heavily debated topics in the U.S. Affective emotion recognition could be hugely influential to understanding the emotional impact of hard news content, yet this relationship is largely still uninvestigated. In consideration of this, we pose the following research questions:

1. How do different multimedia content types (text vs. image) affect the way people emotionally respond to gun-violence-related news?

2. Based on demographic variables, how do people's emotional responses vary to a given multimedia content type?

3. What is the scalability and contribution of a dataset at this intersection?

As a first step towards the exploration of these questions, we extend the Gun Violence Frame Corpus (GVFC) (Liu et al., 2019) along with its corresponding images (Tourni et al., 2021), which contains gun-violence-related news on topics ranging from politics and criminal justice, to social and cultural issues, and more. We present our dataset, BU-NEmo, of gun violence news images and headlines, annotated with elicited emotional response and intensity, along with written explanations to account for any ambiguity in the emotional response. The inclusion of categorical and written annotations can offer an interesting data benchmark on emotional response for both computer vision and natural language processing studies.

This data is, to our knowledge, the first of its kind to:

1. Offer *high-quality affective annotations* on news data from various news sources.

2. Include both *categorical and written* emotional responses on *multimodal news data*.

3. Explore the *differences in emotional response* when given *various modalities of news*.

Additionally, we make our dataset publicly available to facilitate further progress in this area of research[1].

---

[1]The BU-NEmo dataset along with documentation can be found at: https://github.com/Tdrinker/NEmo-dataset

2. Related Works

2.1. Affective Emotion Recognition Background - Psychology & Art

The machine learning task of affective emotion recognition is largely rooted in the backgrounds of art and psychology. The IAPS dataset (Bradley and Lang, 2007) was first composed 15 years ago of roughly 1200 images labeled with their valence and arousal scores for psychological studies. Machajdik and Hanbury (2010) used IAPS as a benchmark to create the ArtPhoto and Abstract paintings datasets, which included artistic works with emotional response annotations. Further, Machajdik and Hanbury (2010) and Zhao et al. (2014) both trained models on these three datasets using hand-crafted features based on art theory principles.

Art is still a large focus for this task, exemplified by the large-scale ArtEmis dataset (Achlioptas et al., 2021), which has around 80k paintings with categorical emotion labels and free-text emotional captions for the purpose of emotion classification and emotion-capable captioning systems. We take inspiration from this SOTA study by adopting their psychologically-backed 8-class emotion categories for annotation, along with collecting written emotional explanations. While their dataset is large-scale, we believe ours is much more timely and high-impact, and due to the nature of artworks, which are driven by the goal of evoking human emotion, the emotional responses and explanations about art pieces are likely incongruent with emotional responses to photojournalism.

2.2. Deep Learning & Affective Emotion Recognition

With the introduction of deep learning technology, various techniques have been implemented on the affective emotion recognition task to improve performance, including creating a multi-level deep representation for images based on their semantics and aesthetics (Rao et al., 2019b; Rao et al., 2019a; Zhang et al., 2020; Li et al., 2019), as well as recognizing and mapping the saliency of "affective regions" in an image during training (Yang et al., 2018; She et al., 2019; He et al., 2019; Rao et al., 2019a; Gao et al., 2020). Other studies have attempted to address the ambiguity of emotion with distribution prediction (Zhao et al., 2017a; Zhao et al., 2017b; Zhao et al., 2020). Rao et al. (2019a) and Gao et al. (2020) have the highest performance among these benchmark studies, with 75-85% accuracy on various emotion-annotated datasets, a full 15-25 point improvement from the foundational non-deep-learning methods in Machajdik and Hanbury (2010) and Zhao et al. (2014).

With the rise in social media platforms and increased availability to visual and textual content, larger-scale single-modal affective emotion recognition datasets have emerged to fuel these performance gains (You et al., 2016; Yang et al., 2017; Balouchian and Foroosh, 2018). All three of these studies scraped images from social applications by searching for emotion-related terms and used these query terms to weakly annotate the data in some form, rather than collecting human annotations. Unlike previous studies, Zhang et al. (2021a) does acknowledge the relationship between various modalities and emotion, but their study also uses data that is weakly-annotated, missing the rich understanding of emotion gained from human-annotation.

2.3. Awareness of Emotional Impact in Social & News Media

As popularity for social media has grown over the years, research its emotional impact on users has also gained spotlight. Using the basis of previous affective emotion recognition studies, Gupta and Yang (2019) and Song et al. (2020) analyzed the impact of emotion on news post shareability and emotionally-triggering image post characteristics, respectively. Li et al. (2016) inspected this from the language side, creating a corpus of news articles annotated with emotion labels to determine which words are most likely to evoke the labeled emotion. Applications such as these have become even more prominent with news coverage of the COVID-19 pandemic (Gupta et al., 2021) and the recent U.S. presidential elections (Thomas and Kovashka, 2019). LD-MAN is perhaps the only work to attempt affective emotion recognition within multimodal news (Guo et al., 2021). While this is the closest study to ours in motivation, the researchers had to create their own dataset due to lack of resources for affective emotion recognition in multimodal news content. The dataset they created was not representative of various news sources, as it only contains articles from the Daily Mail and the Rappler, nor was it given high quality annotations, as only the Rappler data was weakly-annotated by keyword search.

2.4. Emotion in News from a Communication Perspective

Scholars from the communication, media and journalism fields have also investigated various factors that affect people's emotions when consuming political news. Researchers Wanta and Roark (1993) experimented with visuals and text in different conditions within a newspaper setting and discovered that some readers judged the news content solely through accompanying photos. Zillmann et al. (2009) added to this finding on news visuals' affective impact, by stating that the misuse or careless selection of images could lead readers to wrongful perception of the news story.

Other work expanded on these earlier studies, by comparing the impact of news visuals versus texts. Visual elements turned out to be the entry points to a news page (Brantner et al., 2011). When readers had access to both images and headlines, they had a tendency to glance at the images first before reading the headlines. Further, an interesting pattern appeared in Bucher and Schumacher (2006), as news readers used headlines as if they were captions to the images they

saw first. When it comes to specific types of imagery, studies have found that images showing victims, or more broadly, human-interest frames, had particular emotional influence on people (Konstantinidou, 2008). Moreover, a dataset of real news articles for affective emotion recognition is not only necessary, but is highly-impactful for public-interest communication.

3. Method

3.1. Collection of News Images and Headlines

The gun violence news headline and corresponding image dataset used in this design originates from the Gun Violence Frame Corpus (GVFC) (Liu et al., 2019). GVFC contains gun violence news published in 2018 from 21 U.S. news media outlets with a distribution of left, center and right-leaning news organizations. The news articles were pulled from Crimson Hexagon, a social media analytics platform (Hexagon, 2018) by using relevant keywords to U.S. gun violence news, which produced a corpus of gun-violence-related news on topics ranging from politics and criminal justice, to social and cultural issues, impact on the economy and business, and more. This corpus was later extended by Tourni et al. (2021) to include the lead images corresponding to each article. This existing dataset was relevant for our study, given GVFC's accessibility to: 1) images and headlines of news content, 2) a diversity of news sources, and 3) news-framing labels. GVFC ultimately fits our goals of extending emotion annotations to a multimodal news dataset.

3.2. Emotion Annotation Procedure

To understand the emotional response triggered by news samples related to gun violence, we designed a crowdsourcing survey to collect annotations. The annotators were presented with news samples within the context of gun violence in one of three experimental conditions: only the news headlines, only the news images, or both the news headline and image are presented. We will refer to these conditions as *T* (text only), *I* (image only), and *TI* (both text and image) going forward. Do note that the corresponding news headline and image (news set) are from the same news article. In each condition, the annotators were first asked to choose the emotional category they feel among Amusement, Awe, Contentment, Excitement, Fear, Sadness, Anger, and Disgust. These eight categories were used in ArtEmis (Achlioptas et al., 2021), and we decided to follow their annotation schema since these categories are rooted in psychological theory, coming from the prominent study by Mikels et al. (2005). In the next step, annotators were further asked to choose the intensity of their emotional response on a 5-point scale from mild to extreme. Lastly, we asked annotators to describe why they feel the emotion and the intensity of that emotion in written text, to account for ambiguity in emotional response. All three

steps required an answer from the annotators. An example of the interface we used for the three steps of our annotation procedure can be found in figure 1.

Figure 1: Our annotation procedure for a sample of corresponding news image and headline (*TI* condition).

3.3. SONA Platform

We first launched our survey on Amazon Mechanical Turk with annotators who have at least high school qualifications. For each experimental condition (*T*, *I*, and *TI*), we had 130 news sets with each one annotated by 10 different annotators. A significant portion of the annotation we received contained spamming. Multiple attempts were made to improve the quality including an automated script that checks for similarity between the news title and the annotators' written responses to avoid copy pasting and setting a minimum of 5 words to stop annotators from submission with extremely short sentences. However, the spamming in the written part of the responses were not noticeably improved, and thus we also suspect the quality of the categorical annotations, backing up the findings of previous studies that annotating through Amazon MTurk could offer lower quality (Rashtchian et al., 2010).

Due to above limitations and problems occurring in the pilot test of annotations on MTurk, we transferred our experiment to the SONA System that is managed by Boston University's Communication Research Center[2], which is used for annotation purposes at the university. As the participants pool of SONA consists of current undergraduate or graduate students at Boston University, and they only could be granted credits for

---

[2]SONA Website: https://sites.bu.edu/crc/research-resources/sona/

	Number	Percent
Education		
Undergraduate	140	60%
Graduate	95	40%
Citizenship		
United States	125	53%
Foreign	110	47%
Years in the U.S.		
4yr or less	98	42%
5-7yr	29	12%
8+yr	108	46%
Political Leaning		
Left	100	43%
Moderate	115	49%
Right	20	8%
Daily News Consumption		
0hr	9	4%
Less than 1hr	89	38%
1-2hr	87	37%
3-5hr	31	13%
5+hr	19	8%

Table 1: Participant Demographics

their courses or assignments after finishing the research projects' required tasks they participated in, we received high-quality annotation results from this system. Through the SONA survey, we received 15500 total annotation responses from 235 annotators on 840 news sets across the *T*, *I*, and *TI* conditions. A detailed breakdown of our annotator demographics can be found in table 1. We must note that in order to maintain high quality in our data analysis, we only used 10359 of the total annotation responses, coming from 320 news sets and 175 of our annotators. These articles were selected because they have been annotated by at least 10 participants in all three experimental conditions (*T*, *I*, and *TI*). Hence, the other 5141 annotation responses have been temporarily filtered out from our data analysis, but will still be a part of the released dataset.

4. BU-NEmo Dataset Overview

4.1. Annotation Results

In the following figures, we present a basic overview of the resulting data of our annotation experiment. Overall we received diverse responses across all 8 emotion categories, which can be seen in figure 2. There seems to be a much higher frequency of negative emotions, especially in the *I* condition, and in particular, "fear" and "sadness" emotional responses appear the most frequent when images are involved (*I*, *TI*). We tested the independence of our experimental conditions (*T*, *I*, *TI*) on emotional response (categorical) using a Chi-Square test, which confirmed our hypothesis of variable dependence with high confidence and a p-value of $1.9873e{-}14$.

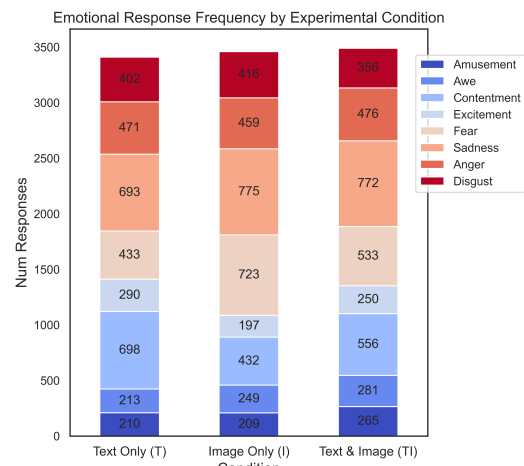The negative emotions (Fear, Sadness, Anger, and Dis-

Figure 2: Overview of categorical emotional responses given in the BU-NEmo dataset by experimental condition (*T*, *I*, *TI*). We have observed more negative emotions in the *I* condition than when text was involved.

	T	I	TI
Avg # Unique Words			
Total Corpus	5.43	5.42	5.68
Amusement	5.39	4.36	5.95
Awe	4.69	4.69	5.55
Contentment	4.88	4.65	5.46
Excitement	4.75	4.81	5.04
Fear	5.68	5.78	5.30
Sadness	5.96	6.12	5.67
Anger	6.00	6.48	6.34
Disgust	6.07	6.46	6.11

Table 2: The average number of unique words used in each written response in our BU-NEmo dataset for each experimental condition (*T*, *I*, *TI*). Calculated for the entire response corpus as well as per categorical emotion chosen.

gust) tended to garner higher intensity scores, but also a wider range of intensity scores than all positive emotions besides Amusement, shown in figure 3. We ran a two-way ANOVA analysis on the dependence between experimental condition, emotion, and intensity scores, and confirmed with high confidence that the condition and emotion have an impact on emotional intensity response, with p-values $8.3474e{-}12$ and $4.2729e{-}46$ respectively.

Table 2 offers an overview of the uniqueness of our written response corpus, showing that there are overall more unique words used in written responses when both images and text are included in the given media (*TI*). The average number of unique words per response for the negative emotions is also significantly higher than for positive emotions in all three conditions, especially for "anger" and "disgust" responses, and even more so when an image is shown (*I*, *TI*).
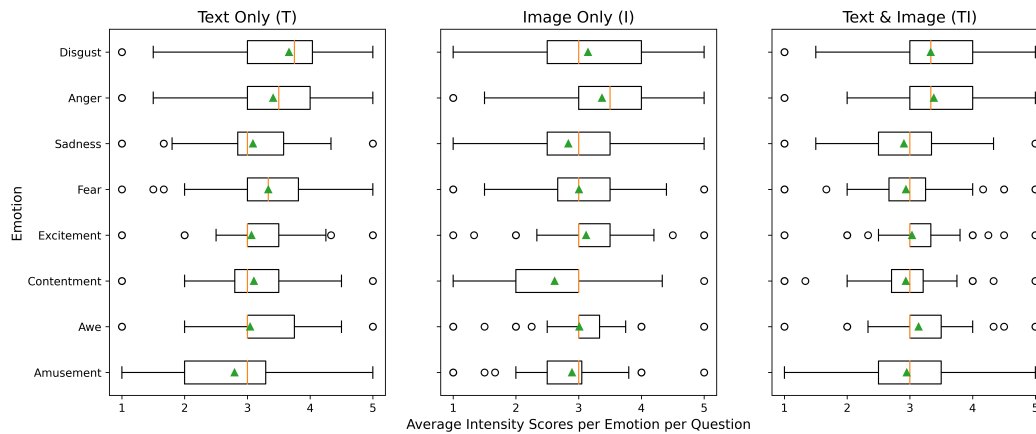
Figure 3: Distribution of emotional intensity responses given by categorical emotional response and experimental condition (*T*, *I*, *TI*). The orange line on each box represents the median, and the green triangle represents the mean. We observed that there are overall higher intensities and larger ranges of intensity scores for the negative emotions.
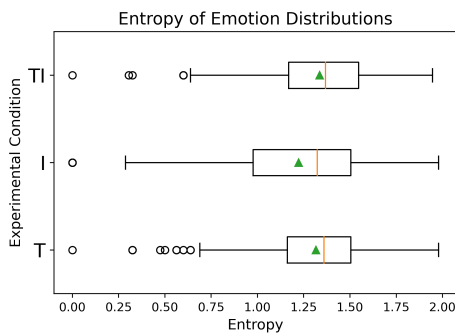
4.2. Response Entropy & Agreement

Figure 4: Entropy over the distribution of emotional responses per experimental condition (*T*, *I*, *TI*). The orange line on each box represents the median, and the green triangle represents the mean. We observe that the entropy is overall lower in the *I* condition, suggesting less consensus when text is a given media type.

In figures 4 and 5, we use entropy to assess the general consensus of annotators in their emotional response. Entropy is a statistical measure that attempts to measure the randomness in a distribution, so in our case, the less random the response distribution is, the more similar the responses are and the smaller the entropy. When examining the consensus of emotional response, entropy is a helpful metric; emotions are not universally experienced, so other metrics that attempt to control for outliers would not make sense. Figure 4 shows the distribution of entropy scores for each question in the *T*, *I*, and *TI* experimental conditions, where the entropy was taken over the emotional response distribution. Overall, entropy scores are more often lower for the image (*I*) condition than when text is introduced

(*T*, *TI*). We verified the significance of this difference in entropy distributions using a one-way ANOVA test, which yielded a p-value of $5.6743\mathrm{e}{-05}$. This pattern may suggest that including text (and thus a specific context) could increase the likelihood of people experiencing the news content with a different emotional response than others.

	Top Emotion	Entropy	Headline
Text Only			
Minimum	Sadness (10/10)	0.0	"The Latest: Victim's family marks anniversary of shooting"
Maximum	Disgust (3/12)	1.9792	"Pittsburgh synagogue-shooting suspect wheeled into courtroom; Trump to visit city Tuesday"
Image Only			
Minimum	Sadness (12/12)	0.0	"California bar massacre leaves another American community reeling after 12 gunned down by Marine veteran"
Maximum	Anger (3/12)	1.9792	"Department of Education considers letting schools buy guns for teachers with federal grant funds"
Text & Image			
Minimum	Sadness (10/10)	0.0	"The Latest: Victim's family marks anniversary of shooting"
Maximum	Awe (4/14)	1.9459	"'Stay in your lane': Doctors fire back at the NRA over guns"

Figure 5: Examples of news with the minimum and maximum entropy in each experimental condition (*T*, *I*, *TI*).

We offer some examples of cases for low and high entropy in figure 5. Interestingly, all three conditions have a sample with minimum entropy (highest agreement) that was voted by all to evoke Sadness. When viewing the headlines corresponding to these samples, it seems as though all three tend to center victims/human interest, offering support for claims that content in these news frames has a particular emotional effect (Konstantinidou, 2008). However, the maximum-entropy case for all three experimental conditions had an agreement rate of 25% among annotators and seemed to correspond to politically-charged headlines. While it has been well documented that emotional responses to con-

2511

tent will not be universal (Roseman et al., 1995; Tsai et al., 2006), these results offer some interesting windows into why that may be the case.
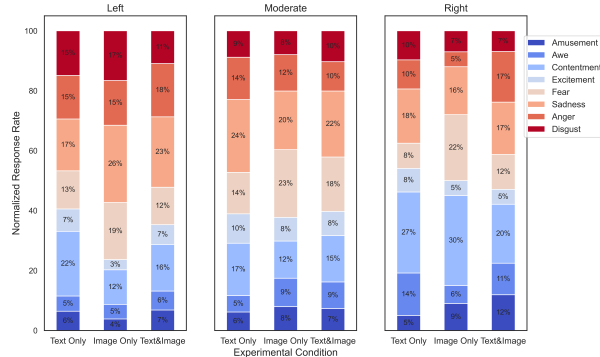
4.3. Demographics & Emotional Response

Figure 6: Emotional responses over each experimental condition (*T*, *I*, *TI*) for each political leaning group. Negative emotions were chosen more frequently overall for left-leaning annotators, while right-leaning annotators favored "contentment" the most frequently.

In figure 6, we show the emotional response rates for each experimental condition, for each political leaning group, taken from our participant demographics. Our participants originally rated their political leaning on a 10-point scale, but we categorized them into left (0-3), moderate (4-6), and right (7-10) for simplicity. It should be noted that the response rates for each emotion are *normalized* with respect to the size of the response group. We opted to present this measure to account for the imbalance in our data related to political leaning, as the amount of right-leaning annotators in our study is significantly smaller than left- and moderate-leaning. This graph clearly shows that the more left-leaning people are, the more they tend to react negatively to the gun-violence news content. While some emotions seem to receive the same amount of response across the board (Amusement and Excitement), other emotions tended to vary greatly depending on annotator political leaning and given media type (Contentment and Fear). In order to verify the statistical significance of this relationship between political leaning, experimental condition, and emotional response, we ran a Chi-Square test, which yielded a p-value of 0.03534. We also examined differences in response related to annotators' daily news media consumption time, but did not find any significant results.

5. Results

In this section, we offer a computational view of the relationship between text and image influence on human emotion, and present a qualitative analysis of the written responses.

5.1. Image & Text Relationship with Emotional Response

To measure how much influence the headline or image of a news article had on its overall emotional response, we analyzed the alignment of responses from the *T* and *I* experimental conditions to the *TI* condition. This alignment was measured at two levels: (1) alignment of categorical emotional response in the given condition to the *TI* condition, and (2) alignment of the intensities of these emotional responses in the given condition to the *TI* condition.

To measure the categorical emotional alignment, we calculated the Negative Emotional Influence (NE Influence) score for the *T* and *I* conditions. This measure calculates the amount of difference between the responses for one condition and the responses of the *TI* condition, giving low scores to the condition with high alignment to the *TI* condition. Firstly, the number of responses (N) for each emotional category e for each condition, as well as this number normalized by the amount of total responses R (NN), are calculated as follows:

$$N_e = \sum_{r \in R} 1_{[r_e = e]} \tag{1}$$

$$NN_e = \frac{N_e}{|R|} \tag{2}$$

where 1 denotes the indicator function. The NE Influence (NEI) for the *T* and *I* experimental conditions is calculated as follows:

$$NEI^T = \sum_e |NN_e^T - NN_e^{TI}| \tag{3}$$

$$NEI^I = \sum_e |NN_e^I - NN_e^{TI}| \tag{4}$$

To measure alignment of emotional intensity scores we calculated the Negative Emotional Intensity Influence (NEI Influence) score for the *T* and *I* conditions. For this, the mean intensity scores (MI) for each emotional category e, normalized by the number of responses for the given emotion, is calculated as follows:

$$MI_e = \sum_{r \in R} \frac{r_i * 1_{[r_e = e]}}{N_e} \tag{5}$$

where r_i is the emotional intensity given by the annotator in response r. The Negative Emotional Intensity Influence (NEII) score for the *T* and *I* conditions is then calculated as follows:

$$NEII^T = \sum_e |MI_e^T - MI_e^{TI}| \tag{6}$$

$$NEII^I = \sum_e |MI_e^I - MI_e^{TI}| \tag{7}$$

Similar to NEI, lower $NEII$ scores reflect higher alignment with *TI* condition responses. We analyzed the resulting NEI scores to determine whether the

headline (T) or image (I) of a given news set had more impact on its emotional response. By comparing NEI^T with NEI^I, we can assess which media stimuli, text or image, likely had more of an influence on the emotional response in the TI condition. If $NEI^T < NEI^I$ then the text is likely to have more influence on emotion than the image for that particular news set, and similarly if $NEI^T > NEI^I$ then the image is likely to have more influence than the text. In our analysis, we have observed that out of 320 news sets, 212 of them have $NEI^T < NEI^I$, 98 of them have $NEI^T > NEI^I$, and 10 of them have $NEI^T = NEI^I$.

For this sample dataset, the analysis shows that text has greater influence on emotional responses than image alone. This also agrees with findings from Zhang et al. (2021a), where authors found that the dominant factor on sentiment between social media images and captions was the caption. However, in our qualitative analysis in the following section, we see that there are some instances when images align more closely with the TI conditions.

5.2. Qualitative Analysis of Written Annotations

We present our qualitative findings from three individual news stories that showed the effect of multimodal content type on the emotion response variations. These examples include the news set, some of its written and categorical responses, and the "Top 10 Keywords" used (top 10 words by frequency in the responses). Before selecting the top 10 words, we removed stop words and special characters, lemmatized words, and removed any mentions of "feel" or "gun violence" from the responses due to redundancy. Not only were there distinctions in responses to the same news presented on different multimodal conditions, but respondents also showed incongruent emotion responses between the chosen categorical emotion and written response.

The first news set (Fig. 7) is a concrete example of when the T condition provided enough information to the news story, allowing for emotional responses similar to this context of the news story. We can see that both T and TI contain positive keywords like "good" which shows people's approval towards the news story. In the I condition, we detect a predominant emotion of fear through keywords like "afraid" and "scary." The I condition contains an image that depicts guns and an American flag, skewing the emotional responses to fear and disgust emotions, while the news content itself was meant to evoke an action being done. Subsequently, when T versus I conditions are compared to the TI condition responses, T closely resembles the TI response results. This shows a potential strength in the text of this news set over the news image.

The second news set (Fig. 8) provides the opposite results from the first, in which the I condition responses provided closer emotional responses to the TI condition than T condition. Additionally, both the TI and I condition responses were reflective of the image's focus on the sadness around the victims and those mourning the victims, compared to the T condition responses which largely consisted of negative sentiments toward the person who caused the massacre. This example shows the potential for the framing of the image to have a more significant effect in people's emotional response to the news, than emotion elicited by the framing of the text.

In the third news set (Fig. 9) that mentions former U.S. president, Donald J. Trump, the T and I conditions contained the frequently chosen emotions such as anger and disgust. However, in the TI condition, many respondents selected contentment, awe, and excitement emotions but wrote their open-ended descriptions with a lot of negative sentiment showing incongruent responses between their chosen emotion and their written responses.

6. Limitations & Future Work

There are certain limitations in our current work. Firstly, the gun violence news dataset from Liu et al. (2019) and Tourni et al. (2021) that we adapt in this work has news framing annotations that we have not utilized. Since only 320 news sets (out of the 1300 sets available in GVFC) have at least 10 annotations in all conditions in our dataset, only a subset of these news framing labels are present. However, we recognize that analysis of news frames and emotions would provide interesting results and hope to include them in the future as we expand our emotion annotations.

Another limitation comes with annotating our dataset on the Boston University SONA System. Due to our population of annotators being Boston University students, our annotators are more alike than if we used a large-scale crowd-sourcing system. This is apparent in the data imbalance along political leaning, with our right-leaning participants accounting for only a small fraction of the group. To account for this, we have normalized data, as explained in section 4.1, when involving political leaning. However, by gathering annotations on a more controlled system, we were able to collect higher quality annotations on our dataset.

Our dataset could be used by others in many applications. According to Xue et al. (2021), fake news will often use highly emotional media to spread quickly, so there is high potential for affective emotion recognition to inform future fake news detection works. Our data could be a great asset to the problem space of affective emotion recognition, especially as a benchmark to measure a model's applicability to real news data. In our own future work, we hope to extend the BU-NEmo dataset with more annotated news data in other news topics, such as climate change. Additionally, we plan to build an ML model to more intricately explore the relationship between modality and emotion.

7. Conclusion

In a world that is increasingly becoming more virtual, the emotional effects of media exposure are becoming

Example 1 | **Top 10 Keywords** | **Response Samples**

Shopify bans sales of certain firearms, accessories

Text
Keywords: 'good', 'access', 'mean', 'decision', 'reduce', 'agree', 'help', 'start', 'actual', 'news'
1. "There is actual action being implemented for gun control." - *Contentment/Happy*
2. "This is saying that change is being made and that companies are preventing people from accessing very dangerous weapons." - *Excitement/Happy* and Excited
3. "It will help to reduce the ownership of guns I believe." - *Contentment/Agree*
4. "Reducing access to buying guns, as part of gun restrictions, makes me feel safer." - *Contentment/Good*

Image
Keywords: 'trigger', 'type', 'people', 'afraid', 'access', 'sale', 'flag', 'scary', 'shelf', 'country'
1. "Guns and the country flag are always associated together, but they shouldn't be" - *Sadness/Sad*
2. "Guns are scary and those are big guns" - *Fear/Fearful*
3. "These types of guns should not just be sold on shelves" - *Disgust/Disgust*
4. "These types of guns are especially dangerous and shouldn't be accessible to citizens under any circumstances." - *Fear/Scared*

Text + Image
Keywords: 'good', 'step', 'right', 'thing', 'control', 'buy', 'safety', 'help', 'small', 'purchase'
1. "This is the right thing to do." -*Amusement/joyful*
2. "It's a good move" - *Contentment/Satisfied*
3. "A small step to help gun control" - *Awe/Awe*
4. "We should look for safety rather than sales of guns." - *Contentment/Contentment*

Figure 7: News article sample where the *T* (headline) condition and *TI* (headline & image) condition aligned most closely on emotional response output.

Example 2 | **Top 10 Keywords** | **Response Samples**

Las Vegas shooting: Memorial events mark 1-year anniversary as massacre motive still shrouded in mystery

Text
Keywords: 'sad', 'people', 'shoot', 'mass', 'event', 'life', 'reason', 'motive', 'know', 'good'
1. "This event was a huge tragedy and I remember hearing about it for the first time. It makes me sad to know that people can actually do terrible things like this and it shows the dark side of humanity. " -*Sadness/sadness*
2. "Person involved in shooting wasn't found and held accountable for what he did" -*Anger/Anger*
3. "The wrecker should be caught." -*Anger/Angry*
4."It's sad that people don't even have to have a motive to create a mass shooting and kill so many innocent people" - *Sadness/Sadness*

Image
Keywords: 'people', 'sad', 'mourn', 'family', 'die', 'person', 'violent', 'lose', 'group', 'prayer'
1. "This implies that someone has died and that made me really sad" -*Sadness/Sadness*
2. "It portrays people that just lost a friend or family member" - *Sadness/Sad*
3. "This appears to be a group of people impacted by a shooting mourning their losses. That makes me sad that they lost people in a violent and avoidable way. " -*Sadness/Sad*
4. "There are so many people who have been affected by gun violence, and who are still hurting" -*Sadness/Sad*

Text + Image
Keywords: 'people', 'sad', 'picture', 'lose', 'die', 'shoot', 'law', 'life', 'closure', 'answer'
1. "It's good to see people coming together at peace and love" -*Contentment/Warm, faith and trust*
2. "Innocent people died and they family have no comfort" -*Sadness/Sadness*
3. "People still don't have the answers or closure for this horrific event." -*Sadness/Sadness*
4. "The people who died in the shooting are not forgotten. But the real problem is not solved." -*Sadness/Sadness*

Figure 8: News article sample where *I* (image) condition and *TI* (headline & image) condition aligned most closely on emotional response output.

Example 3 | **Top 10 Keywords** | **Response Samples**

Trump visits Florida hospital after school shooting

Text
Keywords: 'prevent', 'actually', 'incident', 'trump', 'angry', 'action', 'level', 'f***', 'medium', 'relieve'
1. "How dare he show sympathy to something he is directly contributing to?" - *Disgust/Disgust*
2. "Although he should've visited the hospital, it makes me angry that someone who defines gun rights can also sit by the side of victims of gun violence and still do nothing." - *Anger/Angry*
3. "Trump doesn't actually care and is just doing it for show. Knowing that he won't actually do anything to prevent further incidents like this." - *Amusement/Amused*
4. "Trump only does it out of media coverage and reputation while he does not help with opposing gun violence." - *Disgust/Disgusting*

Image
Keywords: 'trump', 'believe', 'look', 'respect', 'care', 'a******', 'say', 'lot', 's***', 'totally'
1. "Trump looks just like a politician, he totally doesn't care about gun violence in this country." - *Disgust/Disgust*
2."During Trump's presidency, we saw a lot more shootings and I believe it's because those who looked up to Trump believed that he would deem it okay, since he never said anything against gun ownership." - *Disgust/Disgust*
3."The president cannot control gun violence" - *Anger/Angry*
4."Trump is a despicable man, he's a racist, sexist, homophobic piece of **** to be honest." - *Disgust/Disgust*

Text + Image
Keywords: 'good', 'garbage', 'visit', 'matter', 'behavior', 'anger', 'nice', 'option', 'situation', 's***'
1. "Even though he is garbage it was nice he visited, but still a garbage PR move that was only done in that regard. so i guess it sucks but thats not an option" - *Contentment/Contentment*
2. "His behavior won't change the situation" - *Sadness/Sad*
3. "No matter if this is acting, it shows attention from celebrities to gun issues." - *Excitement/Excited*
4. "Besides visiting Florida hospital, he needs to come out with a better gun control regulation" - *Awe/Respect*

Figure 9: News article sample where the *TI* condition yielded incongruent emotional responses from the multiple-choice selection and the written.

more prominent, and AI systems need to be accurately equipped to deal with the intricacies of human emotion. In our work, we demonstrate that the BU-NEmo dataset offers high-quality human emotional response annotations on highly impactful news content centering on the topic of gun violence, as well as open the door for future works in affective emotion recognition and other media-centered tasks to better explore the relationship between news content, its modalities, and the effect it has on news consumers.

8. Ethical Considerations

We would like to acknowledge that we have received permission to use the GVFC dataset (Liu et al., 2019; Tourni et al., 2021), as their data is freely available and has been used only for the purpose of academic research in our study. We plan to make our data available in the same contexts with the publication of this study. Regarding our annotation collection based on their data, we ensure we are not knowingly introducing bias to the data nor inflicting any emotional harm on participants or breaching their confidentiality, for which we have obtained IRB exemption approval (protocol #6280X) from Boston University's Charles River Campus on October 29th, 2021.

9. Acknowledgements

10. Bibliographical References

Achlioptas, P., Ovsjanikov, M., Haydarov, K., Elhoseiny, M., and Guibas, L. (2021). Artemis: Affective language for visual art. *CoRR*, abs/2101.07396.

Balouchian, P. and Foroosh, H. (2018). Context-sensitive single-modality image emotion analysis: A unified architecture from dataset construction to cnn classification. *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 1932–1936.

Bradley, M. M. and Lang, P. J. (2007). The international affective picture system (iaps) in the study of emotion and attention. *J. A. Coan J. J. B. Allen (Eds.), Handbook of emotion elicitation and assessment*, pages 29–46.

Brantner, C., Lobinger, K., and Wetzstein, I. (2011). Effects of visual framing on emotional responses and evaluations of news stories about the gaza conflict 2009. *Journalism & Mass Communication Quarterly*, 88(3):523–540.

Bucher, H.-J. and Schumacher, P. (2006). The relevance of attention for selecting news content. an eye-tracking study on attention patterns in the reception of print and online media. *Communications. The European Journal of Communication Research*, 31:347–368, 09.

Gao, W., Zhang, W., Gao, H., and Zhu, Y. (2020). Visual sentiment analysis via deep multiple clustered instance learning. *Journal of Intelligent Fuzzy Systems*, 39:7217–7231, 11.

Guo, W., Zhang, Y., Cai, X., Meng, L., Yang, J., and Yuan, X. (2021). Ld-man: Layout-driven multimodal attention network for online news sentiment recognition. *IEEE Transactions on Multimedia*, 23:1785–1798.

Gupta, R. and Yang, Y. (2019). Predicting and understanding news social popularity with emotional salience features. pages 139–147, 10.

Gupta, V., Jain, N., Katariya, P., Kumar, A., Mohan, S., Ahmadian, A., and Ferrara, M. (2021). An emotion care model using multimodal textual analysis on covid-19. *Chaos, Solitons Fractals*, 144:110708.

Hasell, A. (2020). Shared emotion: The social amplification of partisan news on twitter. *Digital Journalism*, 9(8):1085–1102.

He, X., Zhang, H., Li, N., Feng, L., and Zheng, F. (2019). A multi-attentive pyramidal model for visual sentiment analysis. pages 1–8, 07.

Konstantinidou, C. (2008). The spectacle of suffering and death: The photographic representation of war in greek newspapers. *Visual Communication - VIS COMMUN*, 7:143–169, 05.

Li, M., Wang, D., Lu, Q., and Long, Y. (2016). Event based emotion classification for news articles. In *PACLIC 30 Proceedings*, Seoul, Republic of Korea, 10. Pacific Asia Conference on Language, Information and Computation.

Li, L., Zhu, X., Hao, Y., Wang, S., Gao, X., and Huang, Q. (2019). A hierarchical cnn-rnn approach for visual emotion classification. *ACM Trans. Multimedia Comput. Commun. Appl.*, 15(3s), dec.

Lin, J. (2021). Amnesty international usa statement for senate judiciary committee hearing on gun violence. *Amnesty International*.

Liu, S., Guo, L., Mays, K., Betke, M., and Wijaya, D. T. (2019). Detecting frames in news headlines and its application to analyzing news framing trends surrounding U.S. gun violence. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 504–514, Hong Kong, China, November. Association for Computational Linguistics.

Machajdik, J. and Hanbury, A. (2010). Affective image classification using features inspired by psychology and art theory. In *Proceedings of the 18th ACM International Conference on Multimedia*, MM '10, page 83–92, New York, NY, USA. Association for Computing Machinery.

Mikels, J., Fredrickson, B., and Larkin, G. e. a. (2005). Emotional category data on images from the international affective picture system. *Behavior Research Methods*, 37:626–630.

Rao, T., Li, X., Zhang, H., and Xu, M. (2019a). Multi-level region-based convolutional neural network for image emotion classification. *Neurocomputing*, 333:429–439.

Rao, T., Xu, M., and Xu, D. (2019b). Learning multilevel deep representations for image emotion classification. *Neural Processing Letters*, 51:2043–2061.

Rashtchian, C., Young, P., Hodosh, M., and Hockenmaier, J. (2010). Collecting image annotations using amazon's mechanical turk. CSLDAMT '10, page 139–147, USA. Association for Computational Linguistics.

Roseman, I. J., Dhawan, N., Rettek, S. I., Naidu, R. K., and Thapa, K. (1995). Cultural differences and cross-cultural similarities in appraisals and emotional responses. *Journal of Cross-Cultural Psychology*, 26(1):23–38.

Schaeffer, K. (2021). Key facts about americans and guns. *Pew Research Center*, Sep.

She, D., Yang, J., Cheng, M.-M., Lai, Y.-K., Rosin, P., and Wang, L. (2019). Wscnet: Weakly supervised coupled networks for visual sentiment classification and detection. *IEEE Transactions on Multimedia*, PP:1–1, 09.

Shearer, E. (2021). 86% of americans get news online from smartphone, computer or tablet. *Pew Research Center*, Jan.

Song, J., Han, K., Lee, D., and Kim, S.-W., (2020). *Understanding Emotions in SNS Images from Posters' Perspectives*, page 450–457. Association for Computing Machinery, New York, NY, USA.

Thomas, C. and Kovashka, A. (2019). Predicting the

politics of an image using webly supervised data. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, number 326, pages 3630–3642, Red Hook, NY, USA. Curran Associates Inc.

Tourni, I., Guo, L., Daryanto, T. H., Zhafransyah, F., Halim, E. E., Jalal, M., Chen, B., Lai, S., Hu, H., Betke, M., Ishwar, P., and Wijaya, D. T. (2021). Detecting frames in news headlines and lead images in U.S. gun violence coverage. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4037–4050, Punta Cana, Dominican Republic, November. Association for Computational Linguistics.

Tsai, J. L., Levenson, R. W., and McCoy, K. (2006). Cultural and temperamental variation in emotional response. *American Psychology Association: Emotion*, 6(3):484–497.

Wanta, W. and Roark, V. (1993). *Cognitive and Affective Responses to Newspaper Photographs [microform] / Wayne Wanta and Virginia Roark*. Distributed by ERIC Clearinghouse [Washington, D.C.].

Westcott, K., Arbanas, J., Downs, K., Arkenberg, C., and Jarvis, D. (2021). Digital media trends, 15th edition. *Deloitte Insights*, May.

Xue, J., Wang, Y., Tian, Y., Li, Y., Shi, L., and Wei, L. (2021). Detecting fake news by exploring the consistency of multimodal data. *Information Processing Management*, 58(5):102610.

Yang, J., Sun, M., and Sun, X. (2017). Learning visual sentiment distributions via augmented conditional probability neural network. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, page 224–230. AAAI Press.

Yang, J., She, D., Sun, M., Cheng, M.-M., Rosin, P. L., and Wang, L. (2018). Visual sentiment prediction based on automatic discovery of affective regions. *Trans. Multi.*, 20(9):2513–2525, sep.

You, Q., Luo, J., Jin, H., and Yang, J. (2016). Building a large scale dataset for image emotion recognition: The fine print and the benchmark. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, page 308–314. AAAI Press.

Zhang, W., He, X., and Lu, W. (2020). Exploring discriminative representations for image emotion recognition with cnns. *IEEE Transactions on Multimedia*, 22(2):515–523.

Zhang, K., Zhu, Y., Zhang, W., and Zhu, Y. (2021a). Cross-modal image sentiment analysis via deep correlation of textual semantic. *Knowledge-Based Systems*, 216:106803.

Zhang, X., Cao, J., Li, X., Sheng, Q., Zhong, L., and Shu, K. (2021b). Mining dual emotion for fake news detection. *Proceedings of the Web Conference 2021*, Apr.

Zhao, S., Gao, Y., Jiang, X., Yao, H., Chua, T.-S., and Sun, X. (2014). Exploring principles-of-art features for image emotion recognition. In *Proceedings of the 22nd ACM International Conference on Multimedia*, MM '14, page 47–56, New York, NY, USA. Association for Computing Machinery.

Zhao, S., Ding, G., Gao, Y., and Han, J. (2017a). Approximating discrete probability distribution of image emotions by multi-modal features fusion. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 4669–4675.

Zhao, S., Yao, H., Gao, Y., Ji, R., and Ding, G. (2017b). Continuous probability distribution prediction of image emotions via multitask shared sparse regression. *Trans. Multi.*, 19(3):632–645, mar.

Zhao, S., Ding, G., Gao, Y., Zhao, X., Tang, Y., Han, J., Yao, H., and Huang, Q. (2020). Discrete probability distribution prediction of image emotions with shared sparse learning. *IEEE Transactions on Affective Computing*, 11:574–587.

Zillmann, D., Gibson, R., and Sargent, S. L. (2009). Effects of photographs in news-magazine reports on issue perception. *Taylor Francis*, Nov.

11.   Language Resource References

Liu, Siyi and Guo, Lei and Mays, Kate and Betke, Margrit and Wijaya, Derry Tanti. (2019). *Detecting Frames in News Headlines and Its Application to Analyzing News Framing Trends Surrounding U.S. Gun Violence*. Association for Computational Linguistics.

Tourni, Isidora and Guo, Lei and Daryanto, Taufiq Husada and Zhafransyah, Fabian and Halim, Edward Edberg and Jalal, Mona and Chen, Boqi and Lai, Sha and Hu, Hengchang and Betke, Margrit and Ishwar, Prakash and Wijaya, Derry Tanti. (2021). *Detecting Frames in News Headlines and Lead Images in U.S. Gun Violence Coverage*. Association for Computational Linguistics.