

# Analysing New York Times Coverage of Sexual Abuse Over Time Through Natural Language Processing

Gyulim Kang<sup>1</sup>, Hope Schroeder<sup>2</sup> and Ian Coppel<sup>1#</sup>

<sup>1</sup>Singapore American School, Singapore

<sup>2</sup>Massachusetts Institute of Technology

#Advisor

## ABSTRACT

Sexual abuse is a prevalent issue in modern society, having acquired prominence during the #MeToo movement in the early twenty-first century. Previous studies note that the media, as a major source of information, plays a significant role in reflecting and shaping societal perceptions of sexual abuse incidents. As a result, analyzing language use in media is crucial in forming an understanding of how inherent biases are conveyed in text. This work aims to investigate the media framing of sexual abuse using NLP techniques such as Valence Arousal Dominance (VAD) analysis and analysis of power and agency scores of verbs. This research investigates the implications of media framing of sexual abuse by incorporating computational modelling and sociolinguistic research. This approach can demonstrate how conscious and unconscious prejudices materialize in text, impacting how specific groups—in this research, victims and perpetrators—are portrayed in the media. Our results show that although the increase in coverage surrounding sexual abuse contributed to an increasingly positive view of female victims from the early 20th century—signaled by an increase in power and agency over time—they were still portrayed as having considerably less power and agency than their male perpetrators. Additionally, girl victims were portrayed with consistently less agency than women victims.

## Introduction

In 2018, across the United States, 81% of women and 43% of men reported experiencing some form of sexual harassment or assault in their lifetime (Kearl, 2018). The issue of sexual abuse is prevalent within modern society, having gained significant attention during the MeToo movement of the early twenty-first century. The MeToo movement was a social movement launched by American activist Tarana Burke with the objective of empowering victims of sexual abuse to publicize allegations of sexual crimes. An accurate representation of sexual abuse in the media is critical because the media is a primary outlet through which the public is exposed to information on cases of sexual abuse and rape. The media can play a positive role in alerting the public about these serious violations of human rights, but it can also at times perpetuate stigma and unconscious biases toward these marginalized groups within the public. Media portrayal in the latter context can have harmful ramifications on protective legislation, social perception, safety, economic security, and more. Therefore, the analysis of language is critical in helping us understand how these biases are expressed in the text through subtle nuances (Entman, 1993).

Previous analyses on the media representation of sexual abuse mainly include content analysis using manual annotations and computational analysis of a smaller scope—De Benedictis' work, for example, analyzes coverage of the #MeToo movement in the New York Times over a span of 6 months. However, a potential downside to these analyses may be that it is difficult to gain a comprehensive picture of the change in coverage of sexual abuse cases over a longer range of time. In order to solve this issue, we conduct a novel analysis of sexual abuse representation in the media over time by focusing on the gender and age of the victim and perpetrator as central markers of information. Gender and age were selected as the focus of this research because information about the gender and age of the victim

and perpetrator is included in the majority of reporting on sexual abuse cases, thus exercising a significant influence on the portrayal of these groups in the media. To this end, we construct a dataset that includes the gender and age of victims and perpetrators as a way to understand better any shifts in framing over time along with the role of power relationships, or the ability of one group to exert control over another within society, in the issue of sexual abuse.

We then use this dataset to study framing differences in New York Times stories of sexual abuse across gender, age, and time. In order to create this dataset, we first take a random sample of articles from the New York Times describing sexual abuse stories, arriving at a total of 23,306 articles. Then, we divide the dataset into three distinct time periods—1970-2000, 2000-2015, 2016-2021—and randomly selected 50-100 articles for each time period to arrive at a total of 196 articles before annotating key information for each article such as the gender and age of the victim and the perpetrator involved in the case depicted. Next, we use NLP techniques including Valence Arousal Dominance (VAD) analysis, Word2Vec, and analysis of power and agency to study emotional valence, word usage, and word embeddings in articles across time, gender, and age (Sap et al., 2017). We apply this framework to a dataset consisting of three distinct time periods of articles in the New York Times, covering a span of over 50 years and analyze how language changed across time on this important topic.

We conduct this analysis to answer three research questions. As the gender and age of the victim and perpetrator within a sexual abuse case are primary areas of interest, we first ask if we observe a difference across the time periods in terms of how cases involving female or male victims are depicted. Next, do we observe a significant difference across the time periods in terms of how cases involving old or young victims are depicted? Finally, how do the agency and power assigned to victims and perpetrators in sexual abuse cases change over time?

Our hypothesis for the first research question is that the framing of cases involving underage victims will use more drastic language than cases involving adult victims; this is because sensationalism of media depictions of young victims tends to portray them in a negative light. Our hypothesis for the second research question is that the framing of cases surrounding female victims will be more negative than cases involving male victims, which may downplay the severity of cases involving male victims. This is due to the increasing attention shown to cases involving female victims, which comprise the majority of cases, compared to those involving male victims. Our hypothesis for the third question is that the change in power and agency scores of victims and perpetrators would be indicative of possible stereotypes, mainly possessing negative connotations that suggest immaturity. In contrast, as years passed and rising global movements such as the #MeToo movement increased general social awareness concerning the importance of sexual abuse-related issues, the words associated with sexual abuse in the media may have taken on a more objective light, deviating from prejudice and instead focusing more on human rights aspects.

The goal of this study is to investigate the implications of media framing of sexual abuse by incorporating computational modeling and sociolinguistic research, holding relevance for a variety of fields. These methods can further reveal how both conscious and unconscious biases are manifested in text, influencing the depiction of oppressed groups in the media. It is critical to gain an understanding of how people and organizations express their outlooks toward specific social groups and issues when constructing our own views regarding these subjects.

## Background

### Sexual Abuse in the United States

Both sexual harassment and sexual assault can be classified as forms of sexual abuse, defined by the American Psychological Association as “unwanted sexual activity, with perpetrators using force, making threats or taking advantage of victims not able to give consent.” As violence against women became a focal issue in the second-wave feminist movement, a robust anti-rape campaign arose in the late 1960s and early 1970s. The campaign grew following the development of the feminist movement and women starting to speak up about their struggles with sexual violence

(McMahon, 2019). The issue of rape had been steadily receiving increased public attention in the media, further contributing to the rise of the campaign.

Moreover, data collected by the FBI revealed that the rate of reported rapes had started to rise rapidly in the 1960s as more females entered the workforce. In the 1970s U.S., rape was the fastest-growing crime (Gornick and Meyer, 2011). Activists recognized the importance of establishing a safe environment for women to gather to share their experiences to decrease the stigma associated with voicing encounters with rape (Gornick and Meyer, 2011). The primary goals of the movement were to advocate for women who have been raped, to educate women on ways to avoid and resist rape, and to urge local institutions—hospitals, law enforcement, and justice systems—to be more responsive to rape victims (Bishop, 2018). The movement experienced numerous victories that included the nationwide establishment of rape crisis centers, organizations advocating the improvement of lives for sexual assault victims, and multiple pieces of legislation (Bishop, 2018).

One of the most notable legislative developments was the Federal Rules of Evidence, also known as the "rape shield" laws, which marked a significant change in how sexual assault cases would be conducted in court. These laws restricted the defendant's ability to investigate the sexual behavior, history, or reputation of the alleged victim to use as evidence to discredit them (Mashel, 2020).

## Media Representations of Sexual Harassment and Abuse

Alongside legislative changes, media representation and perception of these issues also changed over time. In recent years, the media has become the "main conduit of social construction" within society, exercising a widespread influence on the discourse surrounding social issues such as sexual abuse (McCartan et al., 2015). Increased levels of media reporting have contributed to a rise in the visibility of sexual abuse, promoting public awareness (McCartan et al., 2015). In addition, media portrayals of sexual abuse are impactful as they can define what is "deviant" and what is "normal" in a community, affecting how the public identifies sexual abuse (Nair, 2019).

Several studies extensively explore how media affects and constructs public narratives around sexual abuse and child sexual abuse in particular. The abundant existing research on this topic renders a thorough analysis of all its multifaceted findings beyond the scope of this paper. However, three widely discussed trends appear salient to our discussion in this study. First, media practices have driven the shift in the characterization of sexual abuse in the media to emphasize controversy. As Kitzinger asserts, sexual violence is considered to make "good copy." In competition for public attention, stories of sexual abuse have been increasingly sensationalized in response to the public's desire for 'novelty'. Benedict (1992) also states that the style of reporting sexual violence has become "more explicit and lurid over the past few decades," inciting moral panic amongst the audience. Second, studies have highlighted how sexual abuse used to be described as an individual-level problem, whereas it is now portrayed as a societal-level problem with institutional causes. Within sociology, an individual-level issue is defined as being caused by an individual and affecting a small number of individuals. On the other hand, a societal-level issue originates in societal structures and affects many individuals. Third, studies observe the persistence of victim disempowerment and the portrayal of female victims using sexualized terms, although to a lesser degree in more modern time periods (Carter, 1998). For instance, Benedict (1992) states that victims of rape have been stereotyped by the media as either sexualized "vamps" or innocent "virgins."

Returning to the first trend, analyzing the period from 1980 to 1994, Beckett (1996) observes shifts in the framing of child sexual abuse within media discourse through an analysis of four leading magazines: Time, Newsweek, U.S. News and World Report, and People Magazine. In her research, Beckett highlights the transformation in the characterization of sexual abuse over time from a "valence" issue to a "position" issue. Valence issues are defined as issues "characterized by consensus rather than conflict," whereas position issues incite "alternative and highly conflictual responses" (Beckett, 1996; Nelson, 1984). A change in historical context can result in the conversion of an issue characterized by valence qualities into one more adversarial (Beckett, 1996).

Child abuse did not receive significant attention in the media until it was brought to the frontlines in the 1960s when pediatric radiologists recorded damage to children's limbs through X-ray imaging (Beckett, 1996). Reporting laws passed in the 1960s, such as the Child Abuse Prevention and Treatment Act (CAPTA), significantly increased the number of reported sexual abuse cases. The 1970s marked the emergence of writing on rape that framed rape as an issue of "power and control, not mental illness" to hold the perpetrator accountable (Sutherland, 2015). Throughout the 1970s and 1980s, media coverage of sexual abuse and violence gradually increased.

Beckett states the shift in the characterization of sexual abuse from a valence issue to a position issue progressing through this period had several causes, one being media practices. Journalists' increased emphasis on the "dramatic and conflictual" in competition to attract media consumers led to growing attention on false accusations from victims that led to the downfall of the innocent.

Moving onto the second trend, a well-publicized sexual abuse scandal marked the advent of the 21st century: the 2002 Catholic Church scandal. This scandal and the 2012 Sandusky scandal resulted in spikes in coverage from 2002 to 2012, demonstrating the persistence of dramatic focus on egregious abuse cases. Outside of those scandals, there was relatively consistent coverage from 2004-2010. However, Weatherred notes there was a shift among numerous news outlets in the attribution of responsibility for child sexual abuse from an "individual-level problem...to a societal-level problem with institutional culpability" between these two scandals. Focusing on individual-level causes indicates a tendency to view individual behaviors—rather than systemic social issues, which center on institutional policies and legislation—as the source of problems. This tendency results in less attention to the "broader context of societal issues and potential solutions" (Weatherred, 2016).

The #MeToo movement was a transformational social movement that provided victims a platform to speak about their experiences with abuse. The movement, the roots of which originated in 2006, was brought to international attention in 2016 through the Harvard Weinstein scandal when multiple women publicized the abusive behaviors of the American film producer. These cases and many others "lit a spark...but the survivors/victims who came forth against their abuser ignited the movement as a whole" (Tahan, 2021). Regarding the final trend, although numerous studies show the media expanded the #MeToo campaign's visibility by bringing previously unheard stories to the surface of mainstream news, Field states that coverage has been "sympathetic but not necessarily empowering for women" (De Benedictis et al., 2019; Field et al., 2019). Field computes sentiment, power, and agency scores across the corpus to investigate the following questions: "Whom does the media portray as sympathetic?" and "Whom does the media portray as powerful?". Observing broad trends in coverage, Field reports that women were largely absent from the list of highly powered entities, which was instead dominated by men—some of whom were accused of sexual misconduct. Rather, women dominated the list of low-powered entities (Field et al., 2019).

This research analyzes the broader 15-year period of 2000-2015 as one that stands in contrast to the period of extreme legislation change from 1970-2000, the first period analyzed in this study, and the period of significant social change spurred by the #MeToo movement following 2016, the final period analyzed in this study.

## NLP Analysis of Sexual Harassment and Abuse

Previous work has applied NLP to study the framing of social groups and movements in the media. These include an analysis utilizing computational linguistic tactics to investigate the framing of LGBTQ groups in the New York Times over time to understand dehumanization (Mendelsohn et al., 2020). This work pioneers a framework involving specific NLP techniques that can provide valuable insight into linguistic shifts in discussions surrounding marginalized groups. This work informed this research through its use of lexicon-based methods to measure the negative evaluation of LGBTQ groups in the media. These methods could be applied to this research surrounding the analysis of the evaluation of victims and perpetrators involved in sexual abuse cases. These methods included paragraph-level valence analysis, connotation frames of perspective, and word embedding neighbor valence.

First, valence denotes an individual's perspective toward an event or concept and ranges from negative to positive. Mendelsohn et al. utilizes NRC VAD lexicon to calculate paragraph-level valence scores; the lexicon assigns a valence score between zero and one for each verb, with zero representing the lowest valence (most negative emotion) and one representing the highest valence (most positive emotion).

Second, Mendelsohn et al. applies the connotation frames to analyze sentiment directed toward LGBTQ groups, which paragraph-level valence analysis is often too coarse to encompass. This technique involves extracting subject-verb-object tuples within the text using SpaCy dependency parser then using the values within the lexicon to determine the writer's perspective toward the noun phrase.

Third, within the paper, Mendelsohn et al. notes a key drawback of connotation frames analysis that while it can be more comprehensive than paragraph-level valence analysis, it restricts analysis to subject-verb-object tuples. This can omit a large amount of data within the corpus. As a result, Mendelsohn et al. also trains a vector space model to represent each unique word within the data and gain insight into how LGBTQ groups are negatively evaluated. Mendelsohn et al. states the following: "If the vector corresponding to a social group label is located in the semantic embedding space near words with clearly negative evaluations, that group is likely negatively evaluated (and possibly dehumanized) in the text."

Other works center on qualitative analysis, performing a comprehensive content analysis of the dataset based on predetermined markers to gain a deeper understanding of the coverage of social movements such as the Me Too movement in newspapers (De Benedictis et al., 2019). De Benedictis et al. constructed a coding framework that allowed analysis of the movement's in the UK press through aspects such as tone, focus, and the types of solutions it offered for the issues raised by the Me Too movement. Similarly, McGinty et al. (2014) analyzed recent trends in and implications of mental health coverage in the media by devising a 27-item coding instrument, emphasizing the dangers of the public associating serious mental illnesses with gun violence which may lead to the stigmatization of mental illnesses. Along this line of work, Layman (2020) studied how the framing of sexual abuse and rape in the New York Times had changed over time by focusing on elements such as the presence of rape myths and/or stereotypes as well as semantic factors such as the use of euphemistic language or passive voice. Their results observed that rape and sexual assault education and awareness campaigns are more prevalent today than they were in the 1980s, leading to a significant societal shift in perspective about the different aspects that play a role in sexual violence. However, the findings also reported that since the 1980s, victim blaming has increased within The New York Times coverage. Furthermore, the lack of representation of certain victim groups such as male victims seemed to stay consistent throughout (Layman, 2020).

Among the works referenced was also a work that utilized a mixed-method involving both content analysis and textual analysis, believing that limiting the study to one methodology would result in a loss of significant findings (Smith & Pegoraro, 2020). Smith & Pegoraro (2020) identified distinct periods of media coverage with respect to child sexual abuse in the United States and established three research questions, utilizing qualitative content analysis for the first two and questions and textual analysis for the final. The content analysis used a scheme identifying key information within the article such as the month of publication, while the textual analysis conducted thematic and semantic analysis on text-based data by analyzing text in two stages: conceptual extraction (determination of dominant themes) and relational extraction (mapping of relationships of the themes against each other). The findings reported that the focus of coverage was primarily limited criminal justice details of specific incidents rather than contextual information about causes and solutions for CSA. This type of news coverage reinforces the notion that CSA 'just happens' and is a problem too big and complicated to address.

## Data & Methodology

Building on this literature, this study aims to analyze the coverage of sexual abuse cases in the New York Times over time and how social trends may have impacted these trends. To address the primary research questions, we conducted a content analysis of the articles published by the New York Times using the Article Search API. First, we gathered a

broad initial sample by searching with seven keywords ('sexual assault', 'rape', 'sexually assaulting', 'abuse', 'sentenced', 'accused of', 'accuses') to find articles with text that contained at least one of these words. We excluded op-eds to ensure that our sample represented the views of the newspaper instead of individuals. We also excluded articles that were less than 50 words, which were likely to be general factual reports or brief summaries of the details that minimally reflect the publication's perspective. Afterward, we had a total of 23,306 articles. We organized the compiled articles into three time periods and conducted random sampling to reach a total of 196 articles: 1970-2000 (64 articles), 2000-2015 (37 articles), and 2016-2021 (95 articles) based on our analysis of the prominent periods of the feminist movement. From 1970 to 2000, several key pieces of legislation regarding sexual abuse and women's rights were passed, including Title IX (1972), Marital Rape Laws (1993), Megan's Law (1994), and the Violence Against Women Act (1994). Similar to the random sampling method of De Benedictis et al., we randomly chose 50 to 100 articles from each time period (De Benedictis et al., 2019). Like Layman, we choose a small sample to conduct manual labelling. We do not investigate before 1970 due to data availability issues. We then constructed a coding framework that allowed us to analyze the coverage by age and gender labels, which are often central markers of information within sexual abuse cases.

**Table 1.** Subset of annotations of data.

Headline	Text	Victim's Age	Victim's Gender	Perpetrator's Age	Perpetrator's Gender
3 Bronx Patrolmen, On Their Own Time, Seize Rape Suspects	Bronx patrolmen working on their own time arrested two men yesterday...	Y	0	Y	1
U.S. Appeals Court Rejects Death Penalty in a Rape Case as Cruel	A Federal Appeals Court held today that the death penalty for rape...	Y	0	Y	1
Student Charged With 30 Rapes And Robberies in Brooklyn Slum	A 27-year old pharmacy student was arrested last night...	M	0	Y	1
6 Seized in Suffolk in Teenager's Rape	Six Suffolk County youths were arrested today on charges of raping...	T	0	T	1
6 Held in Brooklyn Rape: Two Are 13 and 3 are 15	Six teen-age boys, including two 13-year-olds, attacked and raped a woman...	Y	0	T	1



The letters B, T, Y, and M were used to denote an age less than 10 years, between 10 and 19 years, in the 20's and 30's range, and over 40 years, respectively. We applied this coding framework to the sample of articles in each time period. Table 1 provides an excerpt of the annotations.

In our analysis, we use VAD analysis and analysis of power and agency. VAD stands for valence (the pleasantness of a stimulus), arousal (the intensity of emotion provoked by a stimulus), and dominance (the degree of control exerted by a stimulus). We use VAD analysis to answer our first two research questions and mainly focus on the change over time in the average valence and arousal scores of text portraying sexual abuse cases. This is because dominance is demonstrated by our analysis of power and agency. We conduct this analysis using the NRC VAD lexicon, which has human ratings of valence, arousal, and dominance with scores ranging from zero to one for more than 20,000 English words (Mohammad, 2018). A score of zero represents the lowest valence, arousal, or dominance, while a score of one is the highest possible (Mohammad, 2018). Words with high valence include love and happy, while words with low valence include toxic and nightmare. Similarly, words with high arousal include abduction and exorcism, while words with low arousal include mellow and napping.

To answer our final research question, we analyzed the average agency and power scores of verbs within specific SVO tuples, which we surfaced using the SpaCy dependency parser (Honnibal & Montani, 2017). We then manually approved the tuples that fit our structure of interest: tuples involving both the victim and the perpetrator. For example, we include tuples such as “the perpetrator abused the victim” as both the victim and the perpetrator are included. In contrast, we exclude tuples such as “the policemen arrested the abuser” as only either the victim or the perpetrator is included—in this case, only the perpetrator was included. We used the connotation frames lexicon from Sap et al., which contains power and agency scores for a list of verbs. Power connotations denote the implied authority differentials between the subject (agent) and the object (theme) of the verb. For instance, in the sentence “he begged the attacker to show mercy,” the verb “begged” implies “he” to be in a submissive position compared to “the attacker.” Agency connotation frames imply the thoughts and actions of a person that express their individual power. For instance, the verb “search” in “the victim searched for the perpetrator” implies that “the victim” has a high level of agency. We divided our annotations into eight categories based on gender and age: ‘girl’ victim (female victims under 19), ‘woman’ victim (female victims over 20), ‘girl’ perpetrator (female perpetrators under 19), ‘woman’ perpetrator (female perpetrators over 20), ‘boy’ victim (male victims under 19), ‘man’ victim (male victims over 20), ‘boy’ perpetrator (male perpetrators under 19), and ‘man’ perpetrator (male perpetrators over 20).

We slightly augment the connotation frames for this domain. The lexicon originally did not contain several verbs relevant to our analysis—the excluded verbs mainly described the actions of the perpetrator during their crimes—such as “rape,” “molest,” and “sodomize,” so we added those with obvious power relations according to the original paper’s style. For many of these verbs, both the agency and the power were associated with the agent. In our manual annotations of where these verbs appear in the articles, we did not include verbs where there was negation; an example would be “The man did not take the woman’s life” with the verb “take.”

While VAD analysis and a connotation frames approach can provide valuable insight into broader trends in framing over time, these methods, in the process of quantifying language, provide insufficient information as to how specific words are being used in different contexts. Additionally, the latter limits us to analyzing SVO triples, excluding a significant portion of the available data. To address this issue, we train vector space models to represent the data, thereby capturing the change in specific word use and nuance over time by extracting relations between words in the corpus with our target words.

We use Word2Vec to supplement our results for all three research questions. We begin preprocessing by removing punctuation, lowercasing, tokenizing, removing punctuation, and lemmatizing the data. We then utilize the Word2Vec skip-gram model to generate word embeddings for each sexual abuse-related term within the following list: rape, sexual abuse, sexual assault, sexual harassment, sexual violence, incest, stalking, domestic violence, molest, and sexual exploitation (Mikolov et al., 2013). After training the model, we zero-center and normalize all embeddings to apply a common scale to the data. Finally, we extract the vectors of interest, calculate the cosine similarity with each word in the corpus, and extract the words with the top ten highest cosine similarities.

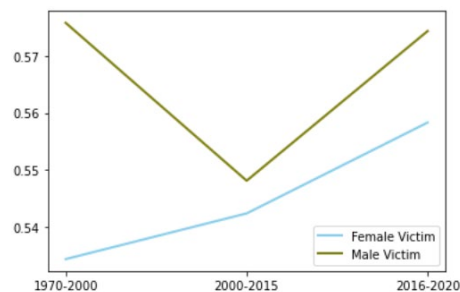
## Results

After completing our analysis, the following results were obtained.

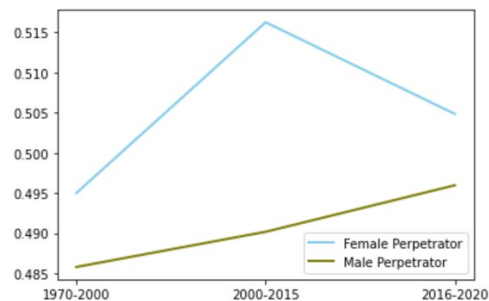
### Research Questions

1. Do we observe a significant difference across the time periods in terms of how cases involving victims of different genders are depicted?
2. Do we observe a significant difference across the time periods in terms of how cases involving victims of different ages are depicted?
3. How do the agency and power assigned to victims and perpetrators in sexual abuse cases change over time?

*Do we observe a significant difference across time periods in terms of how cases involving victims of different genders are depicted?*



**Figure 1.** Change in valence of cases involving female and male victims.



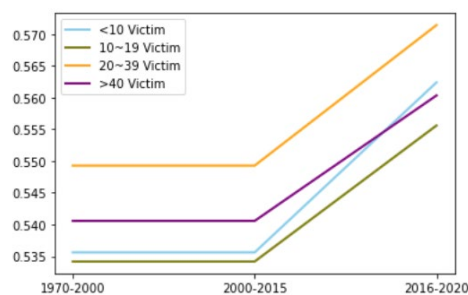
**Figure 2.** Change in arousal of cases involving female and male victims.

To evaluate the media attitude toward victims and perpetrators of different genders, the average valence, arousal, and dominance values of each article were derived through analysis. From Figure 1, it can be observed that the average valence of cases involving female victims continuously increased across the three time periods, suggesting that they were positively evaluated in the New York Times. This may be linked to various feminist legislation and social movements prominent in the 21st century that advocated for female sexual rights. For cases involving male victims, average arousal decreased, suggesting that these cases were less sensationalized over time. This potentially diminishes the fact that males can also be victims of rape, resulting in less attention being brought to their stories. Additionally, the results in Figure 2 suggest that the average arousal increased for cases involving male perpetrators. In view of this result, we can infer these cases were depicted in stronger language over time. This result is in line with the claims of Kitlinger, asserting how the media “[has] sought ever more sensational angles” in depicting sexual abuse cases.



To observe whether the words associated with different sexual abuse-related terms over time indicate differences in the portrayal of victims of different ages, Word2Vec analysis was used. Results show that in the years 1970-2000, the words most highly associated with the list of sexual abuse terms were directly related to the terms themselves, including sodomy, subjected, maltreatment, violence, physical, and denigration (Appendix A). This suggests that reporting during this period centered around the act of sexual abuse rather than those involved. However, this shifted in 2000, when the words “woman” and “children” were included in the top-5 neighbors with highest cosine similarities with 70% of words in the term list. Following the increase in reporting on sexual abuse from the advent of the 21st century, these associations suggest a greater emphasis in media following more on the woman and children sexual abuse victims compared to men; this may also be attributed to the relative prevalence of such cases.

*Do we observe a significant difference across the time periods in terms of how cases involving victims of different ages are depicted?*



**Figure 3.** Change in valence of cases involving victims of all age ranges.

To evaluate the media attitude toward victims and perpetrators of varying age ranges, the average valence, arousal, and dominance values of each article were derived through analysis. From Figure 3, it can be observed that the average valence of cases involving victims of all age ranges increased across the three time periods. Additionally, in cases involving victims of all age ranges, the average arousal increased. The first trend suggests that the overall perception of sexual abuse victims improved as time passed. While sexual abuse was previously viewed as a common occurrence of mild severity, increased media coverage and public attention to sexual abuse in the late 20th century drove a shift in perception that emphasized the violation of basic human rights inherent in sexual abuse. However, with this increase of attention came the rise of media practices that sought to emphasize the dramatic elements of a story in order to capture their audience’s attention and derive profit.

*How do the agency and power assigned to victims and perpetrators in sexual abuse cases change over time?*

**Table 2.** Power scores of female victims over time.

Power Score	Girl Victim	Woman Victim
1970-1999	-0.592	-0.424
2000-2015	0.5	0.909
2016-2021	-0.701	-0.571

We investigate change over time in the power and agency of words used to describe the victims and perpetrators of sexual abuse cases in order to confirm our results derived from the first two research questions. From an age aspect, we can observe from Table 2 that throughout the three time periods, woman victims consistently had more

power than girl victims. This suggests that underage female victims were often portrayed as possessing a lower level of control, or authority, compared to adult female victims. This supports the hypothesis of emphasis surrounding the sensationalization of cases involving younger victims, which may serve to highlight their vulnerability or immaturity as minors. These results reflect the findings of studies such as those of Cohen, who states that events such as sexual abuse are frequently sensationalized and trivialized in the media (Cohen, 1972). Past research further demonstrates how underage victims are portrayed in many articles as either helpless victims or unruly, even wicked people (Franklin and Horwath, 1996). Children's civic responsibility and their status as individuals with rights are given far less attention. Instead, the focus tends to be on children's physical cuteness, novelty appeal, or sensational arrival into an adult-centric world (Goddard and Saunders, 2001).

**Table 3.** Top 5 Closest Neighbours to Word “Molest” During Periods 1970-2000 and 2016-2021

	1970-2000	2016-2021
<b>molest</b>	8yearold	defenseless
	7yearold	powerlessness
	2yearold	subjected
	3yearold	guise
	5weekold	gaslighting

To observe whether the words associated with different sexual-abuse related terms over time demonstrate differences in portrayal of victims of different genders, Word2Vec analysis was used. Table 3 shows that the word “molest” was most highly associated with words denoting age, including 8yearold, 7yearold, 2yearold, 3yearold, and 5weekold. These associations indicate that coverage on sexual abuse may have been more objective, with a greater focus on the ages of victims involved—which range from newborn to toddler—rather than the immorality of the crime. This shifted by 2015, where we no longer see any age-related terms. Instead, the 2015 vector space emphasizes the vulnerability of victims of molestation (suggested by defenseless, powerlessness, and subjected), primarily described as children, in support of the results derived through the power and agency score analysis above.

**Table 4.** Power scores of female victims and male perpetrators over time.

Power Score	Girl Victim	Woman Victim	Boy Perpetrator	Man Perpetrator
1970-1999	-0.592	-0.424	0.632	0.440
2000-2015	0.5	0.909	1	0.958
2016-2021	-0.701	-0.571	0.583	0.588

Primary trends over time included the increase in the agency of female victims. This result is consistent with the results derived through the VAD analysis, where the average valence of articles involving cases of female victims increased. From a gender viewpoint, it can be observed from Table 4 that male perpetrators consistently had more power over female victims. These results support the findings of Field, which showed that despite the goals of the MeToo movement to empower victims, male perpetrators were often depicted as possessing greater individual power to carry out their needs and desires along with having greater power over their victims (Field et al., 2019).

## Conclusion

To conduct our research, we used random sampling techniques to annotate New York Times articles discussing sexual abuse cases; afterward, we utilized VAD analysis along with power and agency scores of verbs within SVO tuples to

evaluate the framing of victims and perpetrators. Our research shows that gender and age have been primary factors influencing the media framing of sexual abuse cases over time. Although the increase in coverage surrounding sexual abuse coupled with legislation promoting victims' rights contributed to an increasingly positive view of female victims from the early 20th century—signalled by an increase in valence and agency over time—they were still portrayed as having considerably less power and agency than their male perpetrators. This provides insight into the extent of the success of the ongoing sexual abuse movement in empowering victims, suggesting that these goals may not be fulfilled yet in media. Finally, our results suggest that media practices that increasingly sensationalized morally repelling cases may have led to the use of more powerful, or intense, language emphasizing the powerlessness of children for cases involving child victims compared to adult victims, resulting in underage female victims having consistently less agency than adult female victims. This research filled the gaps of previous research by investigating a broader range of time periods and focusing on the role of gender and age in the framing of sexual abuse cases through computational analysis.

## Limitations

Several limitations can be found in our method. Within news stories, we found some word senses in the context of sexual abuse that may have different annotator ratings than they would in a general context. One example would include the use of the verb “lock,” which can have different connotations depending on the verb’s object, for example, “locking” a door versus “locking” a person in a room. We use the annotator ratings within the lexicon as a general guide but know that there may be some noise introduced as a result. Future work may involve the development or use of a contextualized lexicon that allows a more nuanced analysis of verb use within the text. Additionally, deeper analyses might result from expanding the size of the dataset and the number of metrics in the coding scheme along with including data beyond the time periods defined in this research.

## Acknowledgements

I would like to thank my mentors and parents for their continued support throughout the research process.

## References

- Allan, S., Branston, G., & Carter, C. (1998). *News, Gender and Power* (1st ed.). Routledge.
- Benedict, H (1992) *Virgin or Vamp: How the Press Covers Sex Crimes*. New York: Oxford University Press.
- Caringella-Macdonald, S. (1991). An Assessment of Rape Reform: Victim and Case Treatment under Michigan’s Model. *International Review of Victimology*, 1(4), 347–361. <https://doi.org/10.1177/026975809100100405>
- Cohen, S. (1972), *Folk devils and moral panics: The creation of the mods and rockers*. McGibbon & Kee, London. <https://doi.org/10.1177/089124167300200308>
- De Benedictis, S., Orgad, S., & Rottenberg, C. (2019). #MeToo, popular feminism and the news : A content analysis of UK newspaper coverage. *European Journal of Cultural Studies*, 22(5–6), 718–738. <https://doi.org/10.1177/1367549419856831>

- Field, A., Bhat, G., & Tsvetkov, Y. (2019). Contextual Affective Analysis: A Case Study of People Portrayals in Online #MeToo Stories. *Proceedings of the International AAAI Conference on Web and Social Media*, 13(01), 158-169. Retrieved from <https://ojs.aaai.org/index.php/ICWSM/article/view/3358>
- Franklin, B. and Horwath, J. (1996), "The media abuse of children: Jake's progress from demonic icon to restored childhood", *Child Abuse Review*, Vol. 15, pp. 310—318.
- Gornick, J. C., & Meyer, D. S. (1998). Changing Political Opportunity: The Anti-Rape Movement and Public Policy. *Journal of Policy History*, 10(4), 367–398. <https://doi.org/10.1017/s0898030600007132>
- Hinze S. W. (2004). 'Am I being over-sensitive?' Women's experience of sexual harassment during medical training. *Health (London, England : 1997)*, 8(1), 101–127. <https://doi.org/10.1177/1363459304038799>
- Honnibal, M., & Montani, I. (2017). *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing*.
- Johnson, L. (1994). [Review of *Virgin or Vamp: How the Press Covers Sex Crimes*, by H. Benedict]. *Gender and Society*, 8(3), 467–469. <http://www.jstor.org/stable/189718>
- Kearl, H. (2018). *The facts behind the #metoo movement: A national study on sexual harassment and assault*. Stop Street Harassment.
- Layman, Katherine E., "The Representation of Rape and Sexual Assault Within News Media" (2020). University Honors Theses. Paper 851. <https://doi.org/10.15760/honors.872>
- Mashel, C. (2020). A Crack in the Armor?: How the Reforms to the New York State Human Rights Law May Expose Weaknesses in Civil Rape Shield Laws. *Fordham Law Review*, 88, 1443.
- McCartan, K. F., Kemshall, H., & Tabachnick, J. (2014). The construction of community understandings of sexual violence: rethinking public, practitioner and policy discourses. *Journal of Sexual Aggression*, 21(1), 100–116. <https://doi.org/10.1080/13552600.2014.945976>
- Mendelsohn, J., Tsvetkov, Y., & Jurafsky, D. (2020). A Framework for the Computational Linguistic Analysis of Dehumanization. *Frontiers in Artificial Intelligence*, 3. <https://doi.org/10.3389/frai.2020.00055>
- Mills, C. W. (1959). *The sociological imagination*. London, United Kingdom: Oxford University Press.
- Mohammad, S. (2018). Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 174–184). Association for Computational Linguistics. doi: 10.18653/v1/P18-1017
- Nair, P. (2019). Child Sexual Abuse and Media: Coverage, Representation and Advocacy. *Institutionalised Children Explorations and Beyond*, 6(1), 38–45. <https://doi.org/10.1177/2349301120190106>
- Patterson, C. J. (1985). [Review of *Making an Issue of Child Abuse: Political Agenda Setting for Social Problems*, by B. J. Nelson]. *American Journal of Education*, 94(1), 125–127. <http://www.jstor.org/stable/1085299>

Seventy-two percent of workers who experienced sex harassment faced retaliation, says new report by NWLC based on Time's up Legal Defense Fund Data. National Women's Law Center. (2022, January 5). Retrieved July 20, 2022, from <https://nwlc.org/press-release/seventy-two-percent-of-workers-who-experienced-sex-harassment-faced-retaliation-says-new-report-by-nwlc-based-on-times-up-legal-defense-fund-data/>

Tahan, Lindsay, "#MeToo Movement: A Sociological Analysis of Media Representations" (2021). Master's Theses, Dissertations, Graduate Research and Major Papers Overview. 339. <https://doi.org/10.28971/1404-OTL65>

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119. <https://doi.org/10.48550/arXiv.1310.4546>

Weatherred, J. L. (2016). Framing Child Sexual Abuse: A Longitudinal Content Analysis of Newspaper and Television Coverage, 2002–2012. *Journal of Child Sexual Abuse*, 26(1), 3–22. <https://doi.org/10.1080/10538712.2016.1257528>