

# Text Mining for a Systematic Literature Review of Sexual Harassment Studies

Dr. Poonam Devdutt, Prof. (Dr.) R.K. Jain, Dr. Naveen Kumar  
Shobhit Institute of Engineering and Technology (Deemed to be University), Meerut  
Email Id- [dr.poonamdevdutt@shobhituniversity.ac.in](mailto:dr.poonamdevdutt@shobhituniversity.ac.in), [Rakesh.jain@shobhituniversity.ac.in](mailto:Rakesh.jain@shobhituniversity.ac.in),  
[naveen.kumar@shobhituniversity.ac.in](mailto:naveen.kumar@shobhituniversity.ac.in)

**ABSTRACT:** *Thousands of research papers have been written on sexual harassment in the twentieth and twenty-first centuries. To summarize the literature on sexual harassment, many review articles have been written. Traditional literature review studies have several disadvantages, such as examining a limited number of articles, being time-consuming and labour-intensive, concentrating on a few subjects, and missing temporal trend analysis. To overcome these constraints, this article uses a combination of computational and qualitative methods to identify key research issues, examine temporal patterns in sexual harassment themes over the last several decades, and suggest future research paths in the field. We gathered 5320 research articles from 1977 to 2020, identified and evaluated sexual harassment issues, and looked at the historical pattern of subjects. According to our results, sexual harassment in the workplace was the most popular study topic, and it was studied in a broad variety of contexts, from schools to military bases. According to our findings, 62.5 percent of the subjects with a significant trend had a rising (hot) temporal trend that would be investigated more in the future years. With text mining, qualitative, and temporal trend analysis techniques, this research provides a bird's eye perspective to better comprehend sexual harassment literature. Researchers, educators, publishers, and legislators may benefit from this study since it provides a comprehensive review of the sexual harassment area.*

**KEYWORDS:** *Literature Review, Sexual Harassment, Text Mining, Temporal Trend Analysis, Topic Model.*

## 1. INTRODUCTION

Sexual Discrimination and Harassment (SDH) is defined by a range of behaviours that range from insulting remarks to sexual and physical assault. Sexual comments (e.g., verbal remarks regarding the size of women's breasts) are the most often reported nonphysical SDH. According to a study, 81 percent of women and 43 percent of men in the United States have experienced sexual harassment at some point in their lives. More than 400,000 Americans over the age of 12 are sexually assaulted or raped each year in the United States, 60,000 children are sexually abused, and 18,900 military members are subjected to unwanted sexual contact. The majority of sexual assault victims (69%) are under the age of 30. According to a study, 21.3 percent of women and 2.6 percent of males in the United States have been subjected to attempted or completed rape at some point in their lives. In terms of where sexual harassment happens, 55 percent occurs at or near the victim's house, 15 percent occurs in an open public place, 12 percent occurs at or near a relative's home, ten percent occurs in an enclosed public area, and eight percent occurs on school grounds. SDH causes physical (e.g., insomnia) and mental health (e.g., depression) problems. SDH has financial ramifications as well. In 2010, more than 12,000 SDH complaints were addressed at a cost of over USD 48 million, resulting in a loss of productivity of approximately USD 22,500 per individual. According to the US Army, turnover accounted for about 67 percent of the overall yearly cost of the harassment recorded. A study found that the lifetime cost of rape per victim is USD 122,461, with a total economic burden of approximately USD 3.1 trillion throughout the victims' lives.

In the twentieth and twenty-first centuries, there has been an explosion of sexual harassment articles produced. Over the past several decades, many review articles have been written to synthesize the literature on sexual harassment. Table 1 lists a few of the important papers and book chapters that have examined sexual harassment studies. These articles were found by

searching Google Scholar for reviews of sexual harassment research (“Sexual harassment” AND “Review”), which contains more research papers than other databases like Web of Science and Scopus. We included reviews that explicitly focused on sexual harassment research as well as some relevant reviews of other broader topics (e.g., work stress) because sexual harassment research is very extensive and can fall under other larger topics (e.g., interpersonal violence and organizational climate). As a result, we included reviews that explicitly focused on sexual harassment research as well as some relevant reviews of other broader topics (e.g., interpersonal violence and organizational climate). We gathered some useful information from this research, such as the number of peer-reviewed publications and the study's subject (s). We may infer that the maximum number of reviewed articles is equal to the total number of cited references for research that did not provide the precise number of examined papers (N/A). For example, in 1991, one research on sexual harassment at work with 68 citations was published, but the number of reviewed articles (#reviewed studies) was not stated. As a result, we may estimate that there will be up to 68 studies evaluated. The number of evaluated papers was not specified in the research, and the total number of citations was 68.

The majority of the reviews focused on workplace sexual harassment and covered the following topics: prevalence and frequency of sexual harassment; legal history and frameworks of sexual harassment; different theoretical perspectives to explain sexual harassment; summaries of precursors at the individual and organizational levels; and consequences of sexual harassment (e.g., retaliation). There were also reviews of sexual harassment interventions and trainings, sexual harassment perceptions and definitions, bullying and peer sexual harassment in youth, sexual harassment in specific settings (e.g., academia) or among specific populations (e.g., women of color, nurses), and reviews of sexual harassment on the Internet. Previous study evaluations are helpful in comprehending many elements of sexual harassment research; nevertheless, these articles and chapters have certain limitations. First, owing to the limitations of conventional literature review papers, a small sample of all relevant papers was chosen from all available relevant publications, mostly from top journals. Second, these publications lack a temporal trend analysis that examines how research subjects evolve over time. Third, the limitations of conventional literature review techniques put still another constraint on studying a restricted number of subjects and often concentrating on a particular topic, such as the frequency and theories of workplace sexual harassment. Fourth, some reviews conducted macrolevel analyses and attempted to synthesize major topics and theories from the literature on workplace sexual harassment; however, these reviews did not cover all aspects of sexual harassment (for example, they did not cover sexual harassment in youth or outside the workplace). Fifth, conventional techniques take a long time to run and cannot handle big datasets[1].

While most literature reviews use standard formats like Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), the current study uses mixed methods to collect and analyse abstracts from a large number of studies in order to fill in the gaps and provide a broader perspective on the sexual harassment literature. To identify key study themes, analyse historical patterns in the subjects over the last several decades, and hint to future potential paths in sexual harassment studies, this article uses both computational and qualitative methods[2].

The following research issues are addressed in this paper:

- What are the major study areas in sexual harassment studies?
- What is the trend of each subject over time?

The following are the benefits of this effort. To our knowledge, this is the first research to look at tens of thousands of sexual harassment papers. Second, the suggested data analysis framework is a versatile method that may be used in a variety of study areas. Finally, the data will be shared, which will allow for additional study and replication of the findings. Fourth, by examining major SDH research issues from 1977 to 2020, this study may provide insight on previous and future sexual harassment research[3].

## 2. REVIEW OF LITERATURE

Paula McDonald in his study discusses about the many women and some men that continue to face workplace sexual harassment (SH) in a variety of professional situations. The goals of this review article are threefold: first, to synthesize the accumulated state of knowledge in the field, with a focus on workplace SH as it relates to management and organizations; second, to evaluate this evidence, highlighting competing perspectives; and third, to canvass areas in need of further investigation. Individual (psychological or legal consciousness) frameworks, sociocultural explanations, and organizational perspectives have all been used to explain workplace sexual violence, but research consistently shows that, like other forms of sexual violence, individuals who experience workplace SH suffer significant psychological, health, and job-related consequences. Despite this, individuals seldom lodge official complaints using internal organizational processes or external authorities. Laws, institutional changes, and policy efforts have been successful in increasing awareness of the issue and shaping employment laws and practices. However, more targeted workplace measures are required to successfully prevent and react to SH[4].

Azy Barak in his study discloses that offline sexual harassment is a well-known, widely-pervasive, well-studied, and well-treated societal issue. Sexual harassment is classified into three categories: gender harassment, unwelcome sexual attention, and sexual coercion, according to a widely recognized paradigm. Sexual harassment behavior is a result of a person's situational qualities, according to theory and research, and it has significant personal and organizational consequences. This article examines the existing knowledge on online sexual harassment, compares it to what has been learnt about offline sexual harassment, identifies particular features of online culture and technology that encourage the behavior, and suggests preventive strategies[5].

Amir Karami in his study discusses about the sheer number of women who have come out about sexual harassment has brought the problem to light. Internet and social media technologies have played a significant role in this transformation. Due to the enormous amount of data posted on these internet technologies, it is difficult to manually evaluate and arrange it; therefore, data and text mining techniques are required. This research offers a predictive framework to gather more than 14,000 sexual harassment complaints on the daily sexism project (ESP) website and identify the space (location) in the reports in order to aid in the battle against sexual harassment. For seven space classes, including the workplace, public space, home, public transportation, school, university, and media, our framework achieves 85.33 percent accuracy. This article also improves the efficiency and efficacy of experiments by combining comparable classes (e.g., school and university) and using a feature selection technique to decrease the amount of characteristics. This enrichment method yields promising results for various classes and characteristics, with accuracy ranging from 86 percent to 93 percent[6].

## 3. DISCUSSION

The corpus and data analysis techniques utilized in this study are described in this section. To study sexual harassment literature including hundreds of research papers, we gathered and cleaned data and used mixed techniques such as topic modeling, topic analysis, and temporal trend analysis. This study's four stages are shown in Figure 1. The sections that follow offer additional information on each stage. For data gathering, Google Scholar does not offer an Application Programming Interface (API). As a result, we'll need to get data from relevant research databases using APIs. We gathered relevant journal and conference abstracts from three major databases, including Web of Science, Scopus, and EBSCO databases, that included the phrase "sexual harassment" in the title or abstract. We discovered abstracts published with their title and keywords after eliminating duplicate entries based on title, abstract, or DOI. Our corpus includes condensed data that represents a broader picture of sexual harassment research articles that addressed a variety of topics. For a variety of reasons, we prefer to look at titles, abstracts, and keywords rather than full-text articles. For starters, although all paper records have a correct title, abstract, and keywords, the full-text format is limited. Second, the title, abstracts, and keywords provide dense, detailed information as well as the most significant results. The full-text format, on the other hand, contains speculative and complicated assertions. Supplementary Material contains the gathered data[7].

## Basic Framework of a Research Proposal



**Figure 1: Research Framework[7].**

With text mining, qualitative, and temporal trend analysis techniques, this research provides a bird's eye perspective to better comprehend sexual harassment literature. The papers covered in this paper were published between 1977 and 2020. Future academics will be able to utilize the findings as a foundation for creating new research ideas and preventive initiatives in the area of sexual harassment. Sexual harassment was investigated in a variety of settings, including the university, school, business, home, online environment, hospital, and military. Researchers were especially interested in workplace sexual harassment. Our findings also indicate that various age groups (e.g., minors), sexualities (e.g., LGBT), races/ethnicities (e.g.,

Latino), and nations were studied in sexual harassment study (e.g., India). Researchers, educators, publishers, and policymakers may find this overview useful in better understanding the broader picture of problems and study subjects, as well as their trends in sexual harassment research. The wide range of studies in the area of sexual harassment demonstrates that harassment's impacts on women and people extend well beyond the conventional workplace. Sexual harassment as a type of sexual violence has been included in research under wider categories, such as bullying in middle school and "elder abuse" in nursing facilities, according to recent studies. Because of the use of broad categories, it's possible that the sexual character of the violence, as well as the link between experiences at different phases of life, may be overlooked. Middle school kids, for example, may be sexually harassed by their classmates, and individuals can be sexually harassed online. We lose their link to the sexual harassment literature if these issues are exclusively examined under the terms "peer and online bullying." In the context of the worldwide COVID-19 epidemic, where millions of individuals have migrated to working from home online, expanding our knowledge of sexual harassment outside the traditional physical workplace is especially important. Thus, a comprehensive examination of subjects that come under the heading of sexual harassment may aid in the formation of linkages between previously researched areas, allowing the discipline to get a greater knowledge of the problem. To synthesize all of the research on sexual harassment published in the last several decades (over 5000 papers) using conventional review techniques would be a monumental task. To the best of the authors' knowledge, no evaluations on sexual harassment have been published that cover such a wide range of topics as this one. The findings are complicated, but they collect and synthesize all studies related to the phrase sexual harassment in the English language between 1970 and 2020. The following are the methodological advantages of this research. As previously stated, the method employed here can evaluate a huge number of articles. Traditional methods are time-consuming and labor-intensive, but this study used a time-saving method. We were also able to detect and forecast hot and cold research topics using the technique we employed here. Finally, the research framework is adaptable and may be used to a variety of study topics. The information provided may assist scholars in comprehending historical changes and trends in research on sexual harassment in different settings. These shifts are most likely the result of historical and contextual variations in society, as well as distinctions across fields. The following discussion places our study's key findings in the perspective of important historical and social movements. Hospital workplace violence, hegemonic masculinity, and gender equality in the workplace were the most popular study subjects in the 1970s. In the United States, legal judgments and definitions of sexual harassment were being established in the 1970s, influenced by Title VII Civil Rights Act 1964 case law and Equal Employment Opportunity Commission policy guidelines (EEOC). Many of the earliest court decisions in the United States that established the terminology and laws around sexual harassment occurred in the 1960s and 1970s. Sexual harassment was defined by the Equal Employment Opportunity Commission (EEOC) in 1980 as "unwelcome sexual approaches or verbal or physical sexual behavior that unreasonably interferes with a person's employment or creates an intimidating or unpleasant work environment". As a result, from a legal standpoint, sexual harassment was initially established in the context of workplace discrimination, and most research in the 1970s and 1980s concentrated on workplace sexual harassment. This tendency is reflected in our findings, with the Workspace category having the most subjects (including Hospital Workplace Violence and Gender Equality in the Workplace). Hegemonic Masculinity, the third most researched subject of the 1970s, is broad and encompasses general study on men's positions of power, including beliefs and behaviors that maintain the status quo. Sexual harassment has been defined as a phenomena that reflects systemic patterns of domination and oppression by people in positions of power (typically males) against individuals who are seen as "other" or "less-than" (often, women. Sexual

harassment is often directed against women since they are seen as others. Similarly, those from less powerful groups are more likely to be sexually harassed, such as males who are considered not masculine enough, LGBTQ people, and people with overlapping identities, such as women of color. As a result, research on how hegemonic masculinity is performed fits with understanding sexual harassment as a way of sustaining historical power relations. Furthermore, throughout the past three decades, Hegemonic Masculinity has been one of the most popular themes (the 2000s, 2010s, and 2020s). The trend analysis for this subject shows no major change (neither hot nor cold); as a result, study in this field has been common and will most likely continue. Gender Equality in the Workplace was one of the subjects studied in the 1980s, indicating a continuing interest in studying sexual harassment in the workplace. Workplace Rules was a second popular subject, suggesting a change away from researching and defining what sexual harassment is to a focus on the policies that are being enacted, perhaps as a result of legislation in the 1960s and 1970s. Finally, Perceptions of Sexual Harassment was the most popular subject in the 1980s, indicating a growing interest in socialpsychology and other disciplines in understanding attitudes and judgements in circumstances including sexual harassment vignettes and scenarios. The prominent fields of study in the 1990s followed a similar pattern, with subjects again linked to the category of Workspace, with Workplace Legal Cases and Workspace Policies being two of the most popular research topics. In a similar vein, research on people's perceptions of sexual harassment has piqued their curiosity. Anita Hill vs. Clarence Thomas was a broadcast case in 1991, and studies show that it altered and influenced how the media and US culture viewed sexual harassment, with Anita Hill becoming a symbol for sexual harassment. While there was considerable interest in legal and legislative elements of sexual harassment throughout this decade, these subjects have since become "cold," suggesting a decline in study in these areas in subsequent decades[8].

Perceptions of Sexual Harassment and Hegemonic Masculinity were prominent subjects in the 2000s, which are consistent with prior decades' study. Youth Bullying and Victimization, on the other hand, was a distinct topic of study this decade, with research focusing on middle and high school students' experiences of peer aggression. Few scholars have looked at school bullying in the United States before to the 1990s, according to academics. Unfortunately, school shootings in the 1990s (e.g., Columbine High School) pushed bullying and other kinds of violence experienced by students in schools into the limelight in America, as well as anti-bullying research and policy implementation in the 2000s. Youth Bullying and Victimization, Hegemonic Masculinity, and Workplace Policies are all hot issues in the 2010s, as they were in earlier decades. Hegemonic Masculinity and Workspace Policies remain the most popular subjects from 2020 until now, indicating consistency in these study fields. Feminism, Media, and Politics, on the other hand, is a relatively recent popular subject. Studies on media depictions of sexual harassment, including the Hill–Thomas hearing, as well as research on the #MeToo movement, are included in this category. The #MeToo and #TimesUp movements reignited conversations about sexual harassment in the United States and across the globe. They sparked fresh lobbying efforts to combat sexual harassment, including many high-profile instances that resulted in penalties for harassers as a result of these campaigns (e.g., Harvey Weinstein being fired and then convicted to 23 years in prison for rape and sexual assault. The #MeToo and #TimesUp movements show how people are utilizing new technologies like social media to connect with other sexual harassment survivors and share their experiences outside of conventional channels (such as the media or workplaces), as well as developing new advocacy tactics. As a result, the subject of Feminism, Media, and Politics demonstrates that scholars are broadening their academic work to incorporate current forums and social movements. In addition to this, another new hot subject, Digital Space, has arisen, indicating that study into the confluence of sexual harassment, social media, and the Internet will likely

continue in the area. Aside from analyzing the most popular subjects in each decade, it's also worth looking into and contextualizing study topics that seem to be growing in popularity. Military Trauma, Healthcare Services, and the Effects of Trauma Exposure are three hot issues in the Health Outcomes area. Researchers have previously looked at the impact of sexual harassment on job-related outcomes such as low employee satisfaction. However, the rise in studies in this area indicates a change in the profession, with a recognition that sexual harassment may be traumatic and that trauma can affect the physical and mental health of people who are exposed to it. Despite the fact that the Workspace category has the most subjects, the majority of them had a cold trend or no change. Medical Field Discrimination and Hospital Workplace Violence are the only two hot issues in this area. As a consequence, our findings reveal a distinct and growing interest in understanding and researching sexual harassment in the medical profession[9].

Domestic violence is a major issue that was formerly classified as a distinct category. While sexual harassment is defined as gendered and sexual violence at work, domestic violence is defined as violence (physical, sexual, and psychological between spouses and is often considered a "private" issue. While these two types of violence are associated with distinct aspects of social life, they are both kinds of interpersonal violence, and there may be a link between them. Both types of violence may be seen as demonstrations and efforts by those in power historically (e.g., males vs. women) to preserve control and conventional power relations. Our findings suggest that study on this subject will almost certainly continue[10].

#### 4. CONCLUSION

This paper offers a broad overview of sexual harassment studies over the last several decades, although it has significant limitations. To begin, just three databases were used to collect data (Web of Science, Scopus, and EBSCO). As a result, it's conceivable that certain subjects were overlooked in our study and are now available in other databases. Second, since we only looked at articles written in English, our review may be missing certain subjects that have been published in other languages. Third, although this study is restricted to research articles including the term "sexual harassment," it does not include all relevant terms. Fourth, although our study focused on research papers, there were other document types worth looking at, such as book chapters, news articles, and opinion editorials, to get a better picture of temporal patterns in sexual harassment research. Fifth, this research examined abstracts owing to restricted access to full-text articles and the time commitment of reading thousands of papers. Sixth, this article offers a breath technique, which may overlook subtle results when compared to conventional literature review approaches. Future research could collect data from other databases, examine non-English manuscripts, look into other keywords and subtopics of the topics identified in this study (e.g., higher education), include other document formats and full-text papers, and analyze topics separately to provide a more in-depth analysis. Researchers may use research databases to acquire full-text versions of articles in a format that is suitable for text mining in 2021, 13, 6589 17 of 24. To offer a more in-depth study, each of the issues (for example, workplace sexual harassment) may be examined individually.

Despite these limitations, this is the only study that we are aware of that attempts to contextualize almost 40 years of sexual harassment research. We were able to catch subjects that were not usually deemed sexual harassment just by looking at the titles of papers since we used a wide approach (e.g., bullying). As a result, the discipline will be able to more critically examine places and people that are suffering sexual harassment but may have been overlooked in previous research. Finally, a greater knowledge of this phenomena will aid in its prevention,

resulting in better and safer workplaces, schools, social spaces, and other areas where sexual harassment is not a concern.

## REFERENCES

- [1] D. Zapf and S. Einarsen, "Bullying in the workplace: Recent trends in research and practice—an introduction," *Eur. J. Work Organ. Psychol.*, 2001, doi: 10.1080/13594320143000807.
- [2] C. Bizios, "#MeToo and sexual harassment training," *New Hampsh. Bus. Rev. ; Concord*, 2017.
- [3] H. J. Hejase, "Sexual Harassment in the Workplace: An Exploratory Study from Lebanon," *J. Manag. Res.*, 2015, doi: 10.5296/jmr.v7i1.6965.
- [4] J. A. Laub, "Assessing the servant organization; Development of the Organizational Leadership Assessment (OLA) model. Dissertation Abstracts International," *Procedia - Soc. Behav. Sci.*, 1999.
- [5] P. A. Frazier, C. C. Cochran, and A. M. Olson, "Social Science Research on Lay Definitions of Sexual Harassment," *J. Soc. Issues*, 1995, doi: 10.1111/j.1540-4560.1995.tb01306.x.
- [6] A. Pina, T. A. Gannon, and B. Saunders, "An overview of the literature on sexual harassment: Perpetrator, theory, and treatment issues," *Aggression and Violent Behavior*. 2009, doi: 10.1016/j.avb.2009.01.002.
- [7] B. A. Gutek and M. P. Koss, "Changed women and changed organizations: Consequences of and coping with sexual harassment," *J. Vocat. Behav.*, 1993, doi: 10.1006/jvbe.1993.1003.
- [8] D. E. Maypole, "Sexual Harassment at Work: A Review of Research and Theory," *Affilia*. 1987, doi: 10.1177/088610998700200104.
- [9] M. V. Studd and U. E. Gattiker, "The evolutionary psychology of sexual harassment in organizations," *Ethol. Sociobiol.*, 1991, doi: 10.1016/0162-3095(91)90021-H.
- [10] A. Barak, "Sexual harassment on the internet," *Social Science Computer Review*. 2005, doi: 10.1177/0894439304271540.