VIOLENCE AGAINST WOMEN: ARTIFICIAL INTELLIGENCE GIVING "VOICE TO VICTIMS"

VIOLÊNCIA CONTRA AS MULHERES: INTELIGÊNCIA ARTIFICIAL DANDO 'VOZ ÀS VÍTIMAS'

Helena Farias Ribeiro[1]

Carla Bonato Marcolin[2]

Cíntia Rodrigues de Oliveira[3]

Abstract: According to World Health Organization (WHO) data, Brazil ranks fifth globally in femicide rates. The 2021 report by the Brazilian Forum for Public Security (FBSP) documented an incidence of one rape every 10 minutes and one femicide every 7 hours. Digital platforms have increased the visibility of victim narratives concerning violence. This study employs artificial intelligence to examine 124 online reports of sexual violence, generating evidence to inform public policy. Following data pre-processing, topic modeling was applied to identify prevalent complaints, perpetrator profiles, and incident locations. The analysis underscores the necessity of careful selection regarding the number of topics to ensure methodological rigor and interpretability of findings.

Keywords: Sexual Violence; Topic Modeling; Social Networks.

Resumo: De acordo com dados da Organização Mundial da Saúde (OMS), o Brasil ocupa o quinto lugar globalmente em taxas de feminicídio. O relatório de 2021 do Fórum Brasileiro de Segurança Pública (FBSP) registrou a ocorrência de um estupro a cada 10 minutos e um feminicídio a cada 7 horas. Plataformas digitais ampliaram a visibilidade de narrativas de vítimas sobre violência. Este estudo emprega inteligência artificial para examinar 124 relatos online de violência sexual, gerando subsídios para políticas públicas. Após pré-processamento dos dados, aplicou-se modelagem de tópicos para identificar queixas prevalentes, perfis de agressores e locais de ocorrência. A análise ressalta a necessidade de seleção criteriosa do número de tópicos para assegurar rigor metodológico e interpretabilidade dos resultados.

Palavras-chave: Violência Sexual; Modelagem de Tópicos; Redes Sociais.

1 Doutoranda – Universidade Federal de Uberlândia; E-mail: helenafariasr@gmail.com; ORCID: https://orcid.org/0000-0002-3884-5856.

2 Doutora em Administração – Universidade Federal do Rio Grande do Sul; Professora – Universidade Federal de Uberlândia; E-mail: carla@ufu.br; ORCID: https://orcid.org/0000-0003-0260-5073.

3 Doutora em Administração – Fundação Getúlio Vargas; Professora – Universidade Federal de Uberlândia; E-mail: cintia@ufu.br; ORCID: https://orcid.org/0000-0001-7999-9002.

Introduction

In 2018, Brazil ranked ninth in the list of the most violent countries, according to the World Health Organization (WHO) (FBSP, 2022). When it comes to violence against women, the situation becomes even more concerning. According to a survey by the Brazilian Forum of Public Security (FBSP), in 2021, Brazil recorded one rape every 10 minutes and one femicide every 7 hours. Additionally, the FBSP noted an increase in crimes against girls and women during the Covid-19 pandemic. Between March 2020 and December 2021, the last month for which data is available, there were 2,451 femicides and 10,398 cases of rape recorded in Brazil (FBSP, 2022). According to the Brazilian Forum of Public Security, the collected data highlights the urgency of determining public policies for the reception, prevention, and confrontation of violence against girls and women in Brazil (FBSP, 2022; Corrêa; Faria, 2021).

According to Dauphin *et al*. (2000), violence against women has been occurring for many years, but it only began to be considered a public problem from 1970 onwards, with the growth of anthropology and feminist perspectives. This was a significant period with intensive intellectual production, during which universities opened up to research groups, supporting and encouraging work on the subject.

Even today, many people consider aggression against women by their spouses justifiable, claiming self-defense (Federal, 2018). This kind of discourse blames the victim, as women are considered guilty for the brutality that happens to them (Borga, 2016). In this sense, women are in a situation of domination, in which those who commit violence exert power and control over the victims. The term power is often associated with the prevention and response to sexual violence, although it is rarely defined or examined (Tatum, 2000).

In 2006, the Maria da Penha Law (Brasil, 2006) was created with the aim of combating domestic and family violence against women. Then, in 2015, the Feminicide Law (Brasil, 2015) was established. In view of this, a crime against life is considered feminicide when it is committed against a victim because she is female. However, these laws are not sufficiently effective, as it is believed that a large proportion of women do not report incidents for fear of their aggressors (Bianchini, 2011). However, there are other reasons pointed out by the victims, such as: believing that violence would not happen again; financial dependence; lack of knowledge about their rights; feeling embarrassed to declare themselves victims; realizing that the aggressor goes unpunished in these situations; and concern for the family (Bianchini, 2011).

With the emergence of the internet and subsequently the intensive use of social media in recent years, reports of victims being shared on social networks have begun to surface, exposing cases of violence suffered or even requests for help. This exposure usually occurs to bring visibility to the report and to try to find ways for the aggressor to be punished by the justice system (Corrêa; Faria, 2021). While human analysis of these reports may involve subjective interpretation, artificial intelligence approaches, such as topic modeling used in this study, can offer a systematic way to process large volumes of data. However, it is important to recognize that AI algorithms can also reflect biases present in their training data and development process. Therefore, this computational approach serves as a complementary tool to help identify patterns and latent topics that can guide public policies and victim protection efforts.

Thus, this article seeks to answer the following question: How can artificial intelligence be used to assist in understanding reports of sexual violence, producing input for public policies? Therefore, its objective is to use artificial intelligence to assist in understanding reports of sexual violence, producing input for public policies.

With the use of artificial intelligence, it is possible to process a large volume of texts in less time. While AI can help reduce some human biases in analysis, it is important to note that algorithms may still reflect biases present in their training data or development process. Thus, problems such as victim blaming, which is exposed in the context of this work, may be mitigated. Over the years, a series of techniques and algorithms capable of summarizing the high-level semantics of content in unstructured text has been developed in the context of text mining, machine learning, natural language processing, and information retrieval (Evangelopoulos; Visinescu, 2012).

Furthermore, when applying topic modeling, it is essential to choose an appropriate number of topics for better interpretation of the output of unstructured models. The number of topics chosen with low held-out likelihood and high semantic coherence will be difficult to interpret and may be considered "junk" topics (Aletras; Stevenson, 2013). Interpretable topics are essential in applications such as visualizing a set of documents (Chaney; Blei, 2012; Newman *et al.*, 2010), where automatically generated topics are used to provide an overview of the documents.

In other words, the output of unstructured models and interpretation depends directly on the choice of the number of topics, as if this is not properly done, an output will be obtained that will not have semantic coherence. Furthermore, the sensitivity of the present context increases

the importance of choosing the number of topics appropriately. Thus, as a methodological contribution, this work discusses the choice of the number of topics, a fundamental parameter of the model's output.

1 Contextualization

Sexual Violence

Studying the history of sexual violence in Western countries reveals various ways in which power is at the root of sexual violence. When Europeans colonized indigenous lands in what is now considered North and South America, they used rape as a tool for power and control (Corrêa; Faria, 2021). Colonizers would rape indigenous peoples as a means of rewarding themselves for the conquest of villages and to keep the indigenous populations living in fear, enabling white colonizers to maintain control over them (Freedman, 2013; Smith, 2012).

Furthermore, slave owners also used rape as a tool for power and control over enslaved individuals. Slaveholders would rape enslaved women to increase their labor supply and, thus, economic power, as the children of enslaved women became the property of the slave owner (Freedman, 2013). Even in the post-emancipation period of the United States of America, white men, particularly those with formal and institutional power such as police officers, used rape as a tool to keep formerly enslaved individuals in a subordinate position (McGuire *et al*., 2010).

These patterns of domination and control persist. Perpetrators of sexual violence, particularly white middle-class men, rarely face consequences or accountability for their actions, and when held accountable for committing sexual violence, these white middle-class men receive lighter sentences than black or working-class men (Alexander *et al*., 2010; Daly; Tonry, 1997). Globally, the World Health Organization estimates that sexual violence affects 12 million people annually, though this statistic likely underrepresents the actual scope of the problem, considering that many cases occur in intrafamilial contexts and do not reach public attention (Baigorria *et al*., 2017).

In Brazil, sexual violence represents a public health problem that reflects similar power dynamics within the national context. Sexual violence accounts for 11.9% of cases of violence against women, with prevalence varying by age group. The rates are higher among adolescents (12 to 17 years old) at 24.3%, compared to young women (18 to 29 years old) at 6.2%, and adults (30 to

59 years old) at 4.3%. Additionally, 71.9% of cases occur in domestic spaces, highlighting the power dynamics within intimate relationships and family structures (Baigorria *et al.*, 2017).

According to the Brazilian Public Security Yearbook, 45,460 cases of rape occurred in Brazil in 2015, averaging 125 cases per day, representing a 10% decline from the previous year. Research conducted in Campinas-SP identified that among women who suffered sexual violence and received hospital care, 52.6% were adults and 47.4% were adolescents, with 69.2% having an unknown perpetrator and 61.5% filing police reports (Baigorria *et al.*, 2017).

The health impacts of sexual violence in Brazil mirror international patterns. Women who experience physical or sexual violence by their partners show more health problems: they have a 16% higher chance of giving birth to low-weight newborns, double the likelihood of abortion, nearly double the probability of depression, and 1.5 times more likelihood of HIV infection compared to women who have not experienced these types of violence (Baigorria *et al.*, 2017).

These Brazilian patterns demonstrate how sexual violence operates as a mechanism of power and control, particularly affecting vulnerable populations and perpetuating cycles of dominance within Brazilian society. The predominance of domestic violence and the targeting of younger women reveal power imbalances that require comprehensive public policy responses tailored to the Brazilian context.

Social networks and smartphones can be allies for women by empowering them to expose, report, and even call for help (Corrêa; Faria, 2021). Pesando (2022), using micro-level data on women and men from demographic and health surveys, including new information on cell phone ownership, analyzed whether women's ownership of cell phones is related to the likelihood of experiencing intimate partner violence (IPV) in 10 low- and middle-income countries. The results showed that women's ownership of cell phones is associated with a reduction of 9% to 12% in the probability of emotional, physical, and sexual violence in the previous 12 months, even after controlling for characteristics representing socioeconomic status, household resources, and local community development. Furthermore, the author asserts that cell phone ownership is positively associated with women's decision-making power within the family and fewer violent attitudes from male partners.

Topic Modeling

Artificial intelligence tools are becoming more prevalent in organizations, enabling greater data processing, analysis, and storage capacity, as well as the creation of alternative data formats. Text mining is one such tool that has contributed to the development of techniques enabling the extraction, summarization, and classification of texts, offering domain expertise through quantitative methods (Marcolin *et al*., 2021).

Topic modeling is a text mining tool, a part of Natural Language Processing (NLP). NLP tools contribute to the search for new insights, that is, to find important data from textual sources that are essential for identifying latent topics in works or selected texts (Bose, 2009). For example, when analyzing hundreds of victim testimonies about sexual violence, topic modeling can automatically identify recurring themes such as "workplace harassment," "domestic violence," or "institutional abuse" without requiring manual categorization of each document.

Topic modeling allows the association of documents based on a set of latent topics with various weights, which define the degree of relevance to a specific topic, as well as reduce the dimensional space. To illustrate, a victim's testimony might be 70% related to the topic "domestic violence," 20% to "psychological trauma," and 10% to "legal proceedings," providing a quantitative understanding of the document's content. Consequently, the result is a clear representation of topics that can be used to study themes addressed in a set of documents, known as a corpus (Aggarwal, 2012).

Topic modeling was initially developed in the 1980s and stemmed from the field of "generative probabilistic modeling." This type of modeling considers that observable variables interact with unobserved or latent parameters in a specific probabilistic relationship, which then produces the data within a dataset. In practical terms, this means that the algorithm assumes that each document is created by combining different topics in varying proportions. For instance, a victim's report might contain words commonly associated with "family violence" (such as "husband," "home," "children") and "physical abuse" (such as "hit," "bruise," "hospital"), allowing the algorithm to identify these underlying themes even when they are not explicitly labeled.

These processes emerged from the demand to describe elements within large data collections without compromising the necessary statistical relationships to perform more direct analyses, such as classification and summarization. Rather than manually reading through thousands of testimonies to identify patterns, researchers can use topic modeling to

automatically discover that, for example, 40% of reports discuss intimate partner violence, 25% involve workplace harassment, and 15% concern institutional abuse, providing a systematic overview of the data's content (Vayansky; Kumar, 2020).

Several current studies in Brazil utilize data mining in the context of violence against women in the country. In the study by Rodrigues *et al.* (2019), methods of topic modeling, sentiment analysis, and news comments were used regarding the attempted femicide of Elaine Caparroz, which occurred in Rio de Janeiro in 2019. Another example of a study employing artificial intelligence is the research by Xue *et al.* (2019), aiming to identify hidden topics in texts about domestic violence on Twitter using the latent dirichlet allocation (LDA) topic extraction technique, commonly used in topic modeling. Data were collected from Twitter, from October 2015 to January 2016, using the keyword "domestic violence," resulting in a corpus of 322,863 English tweets. The results of the study showed patterns identified in user posts, which were then grouped into two relevant themes. The first theme was victimization, indicating that the word "victim" appeared frequently, suggesting a trend on social media platforms towards protecting and supporting victims rather than intervening against perpetrators. The second theme focused on discussing cases of domestic violence that had media coverage, particularly those involving sports figures who committed the crime (Xue *et al.*, 2019).

Unlike previous works, this study utilized longer reports available on the Google search engine, as Twitter has a character limitation, making it difficult for victims to provide a comprehensive report in a single tweet. Additionally, unlike Rodrigues *et al.* (2019), this study used reports provided by Brazilian women, enabling a more comprehensive analysis, as the cited study only examined comments related to the femicide of one victim.

2 Method

This is a documentary study of a quantitative nature, using secondary data collected by Corrêa and Faria (2021). The authors gathered these data using the following keywords in the Google search engine: "Reports of violence against women," "Woman raped uber," "Woman violated," "News woman domestic violence," "Domestic violence report," and "Rape report." No specific time horizon was defined, and there were no criteria for selecting the sources (websites and blogs) to be used. Among these sources are G1, Instagram, Super Interessante, among others.

A total of 124 reports on sexual violence against women in Brazil, collected by Corrêa and Faria (2021), were selected and reprocessed. The data used were in txt format. In the data preprocessing, they were opened in Notepad so they could be concatenated. Accents and punctuation marks were removed from the entire text in Notepad. Subsequently, RStudio was used to remove special characters and stopwords, such as "sao," "nao," "pra," "assim," among others. Stopwords are words that appear in the text but do not have a relevant meaning in defining the topics.

The R language, used to apply topic modeling in this study, is distributed and archived by the CRAN network (https://cran.r-project.org/). It is an open-source software with numerous statistical capabilities, making it a suitable choice for scientific computing. The most important qualities that make R a suitable language are the existence of robust and effective statistical algorithms, access to higher-quality numerical routines, and integrated data visualization tools (Aria; Cuccurullo, 2017).

After data preprocessing, the LDA technique was applied using the STM (Structural Topic Modeling) package to perform topic modeling. The number of topics was defined by exploring the dataset to calculate four metrics: held-out likelihood, semantic coherence, residual, and lower bound. In this sense, to identify the model that reproduces more semantically coherent and distinct topics, the four metrics were calculated in a range from 3 to 10 topics (Benoit, 2018).

3 Results

For the initial exploration of the 124 texts, a word cloud (Figure 1) was generated with the most frequent words in the reports, with a minimum frequency of 10 occurrences per word and a maximum of 50 words. It is possible to observe that the most prominent words are "home" (casa, in Portuguese), "mother" (mãe, in Portuguese), and "everything" (tudo, in Portuguese). This result was generated to compare it with the latent topics that will be identified in Table 1.
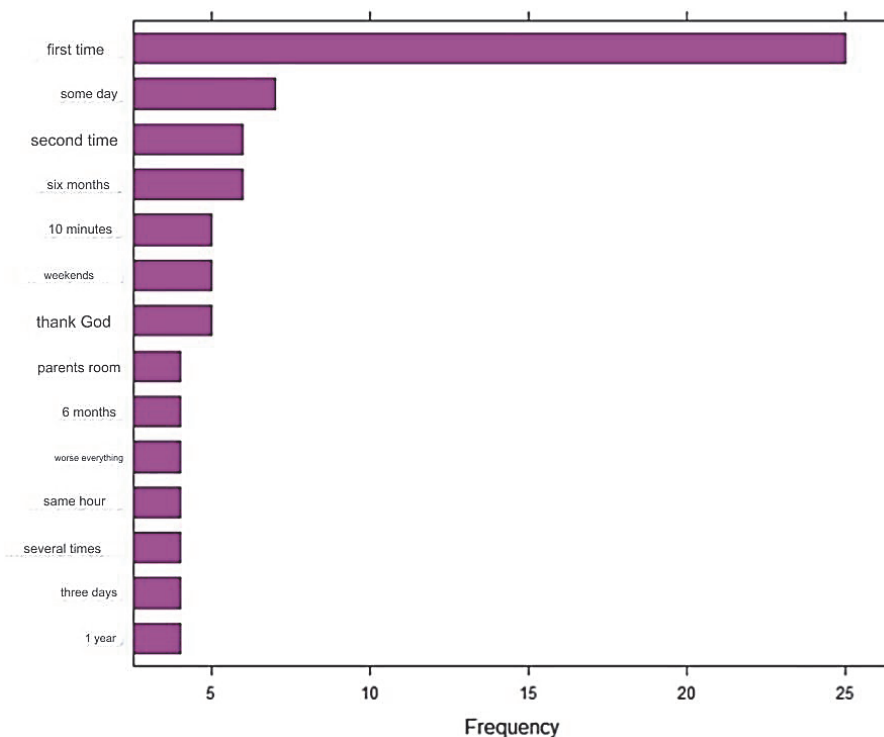
Figure 1- Wordcloud
Source: Authors' own elaboration (2025).

Another exploratory analysis conducted was the identification of grammatical classes, also known as Part-Of-Speech tagging (POS) technique. Figure 2 presents the most frequent noun-verb pairs found in the reports. It is possible to notice that many reports tell of the first times these women were abused, which is why "first time" appears first. Second comes "certain day," which is used in several reports. In third place appears "second time," indicating that many victims suffered abuse more than once, as in the following excerpts: "I was abused, touched violently for the first time. The second time I think I was still around 5 years old when I went on vacation at my grandmother's house..."; "The second time I was sexually abused was when I was 6-7 years old by the brother of a friend of mine who at the time was a teenager, around 16-17 years old..."

Figure 2 - Most frequent word pairs (noun-verb)
Source: Authors' own elaboration (2025).

For topic modeling analysis, careful selection of the number of topics is necessary. To do this, different measures of model quality were evaluated. In particular, semantic coherence, which assesses the co-occurrence of words within documents to ensure that the selected keywords belong to a single concept, thus maintaining interpretability and topic quality (Mimno *et al.*, 2011). The five-topic model was selected for its high semantic coherence and held-out likelihood, with low lower bound and residual as shown in Figure 3 (Blei *et al.*, 2003).
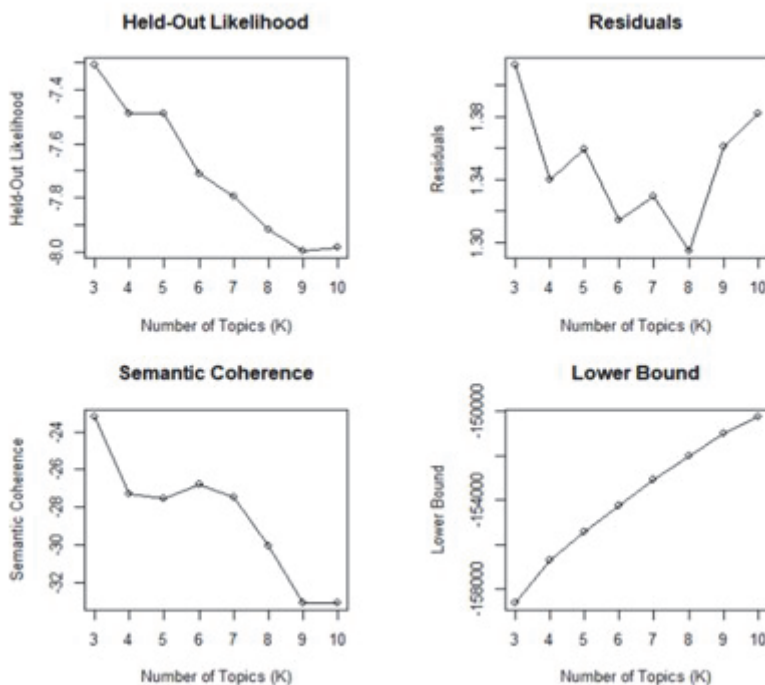
Diagnóstico de valores por números de tópicos

Figure 3 - Diagnosis of values by number of topics.
Source: Authors' own elaboration (2025).

After choosing five topics, the model was run again, and the results are shown in Figure 4. It can be observed that Topic 2 presents the highest proportion in the reports, containing the words "year", "father" and "never". This suggests that the attacks occurred over extended periods, such as a year, and may have been carried out within the victim's family circle, indicating that the aggressor could be the father. The word "never" appears in passages where victims state that they never anticipated suffering abuse, as in the following examples: "although several friends had difficult moments, I never thought it would happen to me..."; "Once he squeezed my arm tightly while asking me to repeat that I was only his. But I knew he wouldn't hit me, he had never laid a hand on a woman to do that...".

Some excerpts in which the word "father" appears include: "When I was 14 years old I discovered I was a lesbian. My parents are separated. My father took me traveling with him one weekend. And there the first abuse happened. I didn't understand what was happening. I was paralyzed with fear.

I felt disgusted and cried but he continued…", "I had no reaction, that man who I considered my father was proposing to photograph me naked… what do you mean? He talked nonstop about the pros of negotiation, I no longer listened. I couldn't believe my eyes, my ears, was I really there?…", "If we didn't do it, they would tell my father that I liked being naked for them. As my father always said that it was my fault for everything, I started to accept these things – his thinking was that women serve to fulfill men's wishes…". These data highlight the phenomenon of sexual violence against children, not only regarding the fact that most victims are attacked within their own family circle (since the word "home" appears frequently in topics 1 and 5 of Figure 2 and in all topics in Table 1), but also concerning the type of relationship between victim and aggressor.

According to Furniss (1993), the figure of the father is identified as the most frequent perpetrator in sexual aggression against children. According to Ferreira (2005), the father is the primary perpetrator of sexual violence against children, followed by a high incidence of other family members, such as the stepfather.
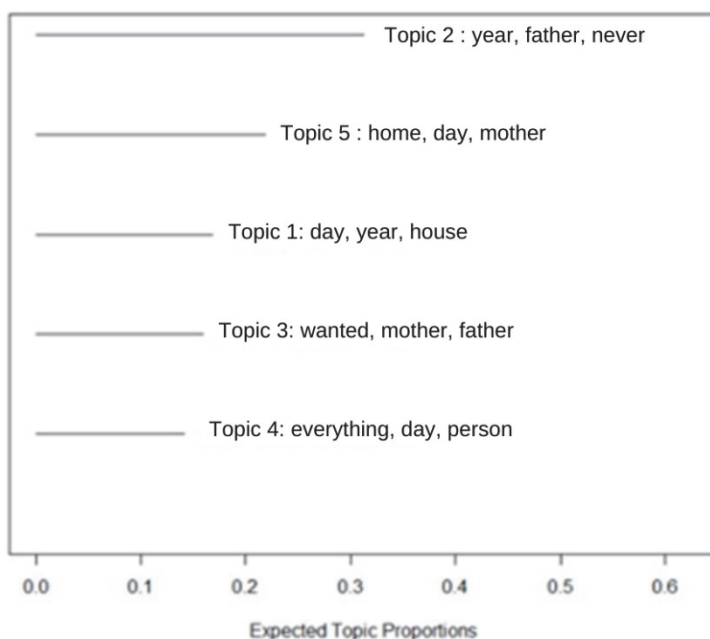

Figure 4 - Most relevant topics
Source: Authors' own elaboration (2025).

In Topic 5, which appears second in proportion in Figure 4, the words "house", "day" and "mother" appear. It is possible to relate these to reports from victims who suffered domestic violence at home, where the mother often appears unaware of what was happening, as in this excerpt from a report: "I looked at my mother who was sleeping in the bed opposite me and I didn't know what to do, that's when he took me to the living room and stayed there with me until he felt satisfied." This reinforces the power relationship that this father had over his daughter, which kept her unable to report the incident or even tell her mother. The word "mother" also appears in many reports where children describe the sexual violence they witnessed in their father's home against their mother. It is important to note that the sexual violence described in the reports encompasses not only penetration during sexual acts, but also inappropriate touching, abusive language, and other forms of assault.

Subsequently, seven columns were selected for each of these topics, that is, seven words for each latent topic that are presented in the results. Therefore, the results display five rows of topics, with seven words each (Table 1).

Table 1 - Most relevant latent topics

Topic 1	Day, Year, House, All, Mother, Still, Fear
Topic 2	Year, Father, Never, House, Mother, Always, Today
Topic 3	Wanted, Mother, Father, House, Day, Said, With me
Topic 4	All, Day, Person, Life, Whole, Always, House
Topic 5	House, Day, Mother, Nothing, Friend, Said, Car

Source: Authors' own elaboration (2025).

Observing the topics in Table 1, it is possible to notice similarities between them. In Topic 1, these words together suggest that the victim suffered violence daily throughout the year (this could refer to the mother, who appears as a word, or even the children) and that the victim still experiences fear. In Topic 2, the pattern is similar to the first topic, indicating that the victim suffered violence at home, which could involve the mother or the children, the father appears as a possible aggressor, and the victim feels that they would never be able to escape that situation. This becomes clear in this excerpt from a collected report: "My mother tried to stop her and was pushed and almost got beaten too. I, in my idealistic feminism, told him that if he hit me I would report him for domestic violence. He mockingly replied: then call the police, call 190 to come and arrest me…".

In the other topics, it is evident that the attacks occurred at home, possibly daily, as the word "all" appears in Topic 4, and the victims' perception is that

it would never end. The word "friend" appears in Topic 5, as in some reports the victims ask a friend for help to leave the house and escape the violent situation, or suffered violence at their friend's house, by a family member, as in the following excerpt: "He was the father of a friend of mine. After class, I went to her house to study like we always did. He actually worked at the school I attended. And, at a certain time of the day, he said that each one should stop studying and start helping to clean the house. He asked me to hang out the clothes. I ended up standing against the wall, where the clothesline was attached, to hang the clothes, and that's when he came and rubbed himself against me while my back was turned…".

The results of the topics corroborate the literature review, which indicates that although there are laws addressing feminicide and domestic violence, Brazil needs to make significant progress in combating this type of crime, as the aggressor is often not held accountable and is soon free to commit violence again, putting the lives of the victims who report it at risk.

Considering the importance of evaluating the stability of models based on artificial intelligence, an inappropriate number of topics was considered as shown in Figure 3. It can be observed that the number 7 topics represents lower held-out probability and greater semantic coherence, indicating lower overall model quality (Roberts *et al.*, 2014). Figure 5 contains the results of the main topics, considering the 7.
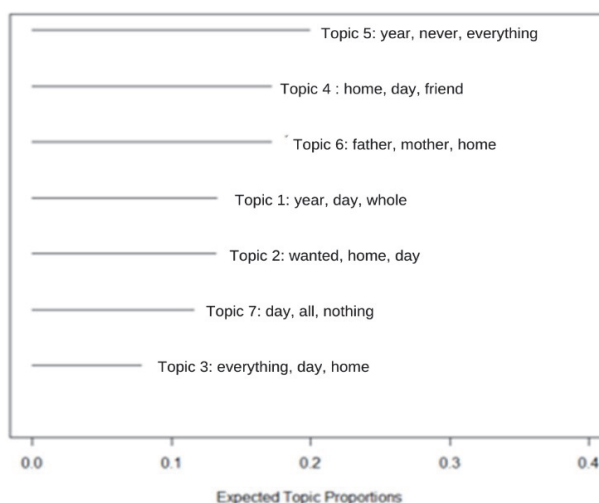

Figure 5. Most relevant topics, considering 7 topics.

Source: Authors' own elaboration (2025).

To allow a comparison between the interpretability of the two models, the seven top words considering the seven topics were selected (Table 2). It can be observed that it becomes more difficult to interpret the topics from Table 2 than from Table 1, due to the low held‑out likelihood. This indicates that the words do not have a satisfactory connection among themselves to represent the studied reports.

Table 2 - Latent topics for testing with 7 topics

Topic 1	year, day, all, time, fear, still, everything
Topic 2	wanted, house, day, mother, all, said, father
Topic 3	all, day, house, life, time, person, thing
Topic 4	house, day, friend, mother, said, nothing, whole
Topic 5	year, never, all, always, whole, day, mother
Topic 6	father, mother, house, year, day, stayed, because
Topic 7	day, all, nothing, friend, woman, happened, house

Source: Authors' own elaboration (2025).

Furthermore, observing the main topics from Figure 5, it is noticeable that the primary topic contains words like "year", "never", "all", which do not represent anything significant. On the other hand, in Figure 4, the words "year", "father", "never" appear together, representing part of the reports, as analyzed in Table 1.

Conclusion

The study successfully addressed the research question: How to use artificial intelligence to assist in understanding reports of sexual violence, providing input for public policies? In this context, the objective of using artificial intelligence to aid in the understanding of reports of sexual violence, generating input for public policies, was achieved. Based on the results of the study, it is evident how important it is to provide sexual education to children from a young age. How can they feel safe if the abuse occurs within their own homes?

In this regard, schools play a key role, even taking on the responsibility of receiving reports of violence from children. Therefore, it is essential to train and capacitate teachers, educators, and pedagogues working in schools, nurseries, and other institutions, to teach children how to protect themselves, even from their own fathers, and to recognize what constitutes aggression and abuse. Hence, it is important for children to understand their intimate parts and that nobody (even if it is their father or another relative) should touch them unless it is for hygiene purposes. Moreover, schools need to be trained to handle cases of suspected sexual violence or abuse, as knowing how to respond correctly after identifying suspicions of abuse is the first step in effectively making a report.

With sexual education in schools, it is also possible for children to report the abuse suffered by their mothers. Furthermore, it is essential to review laws regarding sexual violence, as victims will not feel safe to report unless the perpetrators are fairly punished. Brazil recorded one rape every 10 minutes and one femicide every 7 hours in 2021 (FBSP, 2022).

Additionally, this study highlighted the importance of selecting the appropriate number of topics to ensure a good interpretation of the reports. According to Tables 1 and 2, and Figures 3 and 5, it was evident that considering the number of topics as 5, instead of 7, allowed for a better interpretation of the studied reports, particularly in such a sensitive topic.

The limitation of this study lies in the use of secondary data and the limited number of keywords used in the Google search. Therefore, it would be valuable to replicate the methodology with a more robust database, such as a case study, using other keywords, to compare the results and guide public policies that genuinely protect women from sexual violence.

References

AGGARWAL, C. C. **Mining Text Data**. Londres: Springer International Publishing Switzerland, 2012.

ALETRAS, N.; STEVENSON, M. Evaluating topic coherence using distributional semantics. *In*: INTERNATIONAL CONFERENCE ON COMPUTATIONAL SEMANTICS, 10., 2013, Potsdam. **Proceedings** [...]. Potsdam, 2013. p. 13-22.

ARIA, M.; CUCCURULLO, C. Bibliometrix: An R-tool for comprehensive science mapping analysis. **Journal of Informetrics**, v. 11, n. 4, p. 959-975, 2017. https://doi.org/10.1016/j.joi.2017.08.007.

BAIGORRIA, J. *et al*. Prevalência e fatores associados da violência sexual contra a mulher: revisão sistemática. **Revista de Salud Pública**, v. 19, p. 818-826, 2017.

BENOIT, K. **stm.R**. 2018. Disponível em: https://github.com/bstewart/stm/blob/525b00ced93eac13881e63fd4ccd25771682a2fd/R/stm.R#L22. Acesso em: 27 out. 2022.

BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent Dirichlet Allocation. *In*: International Conference on Machine Learning, 3., 2003. **Proceedings** [...]. 2003. p. 993-1022.

BORGES, M. "Eu não vou te matar": um retrato da violência doméstica. **Instituto Humanitas Unisinos**, 2016. Disponível em: http://www.ihu.unisinos.br/noticias/550570-eu-nao-vou-te-matar-um-retrato-da-violencia-domestica. Acesso em: 26 out. 2022.

CORRÊA, I. T.; FARIA, E. R. An analysis of violence against women based on victims' reports. *In*: Brazilian Symposium on Information Systems, 17., 2021, Curitiba. **Anais** [...]. p. 1-8. https://doi.org/10.1145/3466933.3466968.

DAUPHIN, C. *et al*. A história das mulheres. cultura e poder das mulheres: ensaio de historiografia. **Revista do Núcleo Transdisciplinar de Estudos de Gênero**, v. 2, n. 1, p. 7-30, 2000. Disponível em: http://www.marilia.unesp.br/Home/Pesquisa/cultgen/Documentos/historia_das_mulheres_nuteg.pdf. Acesso em: 28 out. 2022.

EVANGELOPOULOS, N. E.; VISINESCU, L. L. Text-mining the voice of the people. **Communications of the ACM**, v. 55, n. 2, p. 62-69, 2012. https://doi.org/10.1145/2076450.2076467.

FEDERAL, Senado. **Panorama da violência contra as mulheres no Brasil**. 2018. Disponível em: http://www.senado.gov.br/institucional/datasenado/omv/indicadores/relatorios/BR-2018.pdf. Acesso em: 27 out. 2022.

FERREIRA, A. L. Acompanhamento de crianças vítimas de violência: desafios para o pediatra. **Jornal de Pediatria**, v. 81, p. s173-s180, 2005.

FÓRUM BRASILEIRO DE SEGURANÇA PÚBLICA (FBSP). **Anuário Brasileiro de Segurança Pública**. 2022. Disponível em: https://forumseguranca.org.br/wp-content/uploads/2022/07/10-anuario-2022-feminicidios-caem-mas-outras-formas-de-violencia-contra-meninas-e-mulheres-crescem-em-2021.pdf. Acesso em: 25 out. 2022.

FREEDMAN, E. B. **Redefining Rape**. Cambridge: Harvard University Press, 2013.

FURNISS, T. Abuso sexual da criança: uma abordagem multidisciplinar, manejo, terapia e intervenção legal. *In*: **Abuso sexual da criança:** uma abordagem multidisciplinar, manejo, terapia e intervenção legal. 1993. p. 337-337.

MARCOLIN, C. B. *et al*. Do explanatory notes explain? Risk management disclosure analysis using mineração de textos. **Revista Contemporânea de Contabilidade**, v. 18, n. 48, p. 1-18, 2021. https://doi.org/10.5007/2175-8069.2021.e78104.

MCGUIRE, K.; STEWART, A. J.; CURTIN, N. Becoming feminist activists: Comparing narratives. **Feminist Studies**, v. 36, n. 1, p. 99-125, 2010.

MIMNO, D. *et al*. Optimizing semantic coherence in topic models. *In*: Conference on Empirical Methods in Natural Language Processing, 2011. **Proceedings** [...]. 2011. p. 262-272.

MULHERES, Secretaria Especial de Políticas para as. **Lei Maria da Penha**. Lei nº 11.340, de 7 de agosto de 2006. 2006. Disponível em: https://presrepublica.

jusbrasil.com.br/legislacao/95552/lei-maria-da-penha-lei-11340-06. Acesso em: 26 out. 2022.

NEWMAN, D. *et al.* Visualizing search results and document collections using topic maps. **Web Semantics:** Science, Services and Agents on the World Wide Web, v. 8, n. 2-3, p. 169-175, 2010. https://doi.org/10.1016/j.websem.2010.03.005.

PESANDO, L. M. Safer if connected? Mobile technology and intimate partner violence. **Demography**, v. 59, n. 2, p. 653-684, 2022. https://doi.org/10.1215/00703370-9774978.

ROBERTS, M. E.; STEWART, B. M.; TINGLEY, D. Stm: R Package for Structural Topic Models. **Journal of Statistical Software**, [s.d.].

RODRIGUES, L. D. F.; DA SILVA JUNIOR, J. L. F.; LOBATO, F. M. F. A culpa é dela! E isso o que dizem nos comentários das notícias sobre a tentativa de feminicídio de Elaine Caparroz. *In*: Brazilian Workshop on Social Network Analysis and Mining, 8., 2019, São Paulo. **Anais** [...]. p. 47-58. https://doi.org/10.5753/brasnam.2019.6547.

ROUSSEFF, D. **Lei nº 13.104,** de 9 de março de 2015. Lei do feminicídio. Diário Oficial da União, Brasília, DF, 9 mar. 2015. Seção 1, p. 1.

SAITO, M. I.; LEAL, M. M. Educação sexual na escola. **Pediatria**, v. 22, n. 1, p. 44-48, 2000.

SMITH, A. Indigeneity, settler colonialism, white supremacy. *In*: **Racial Formation in the Twenty-First Century**. 2012. p. 66.

TATUM, B. D. The complexity of identity: "Who am I?". *In*: **Readings for Diversity and Social Justice**. v. 2, 2000. p. 5-8.

VAYANSKY, I.; KUMAR, S. A. P. A review of modelagem de tópicos methods. **Information Systems**, v. 94, p. 101582, 2020.

XUE, J.; CHEN, J.; GELLES, R. Using data mining techniques to examine domestic violence topics on Twitter. **Violence and Gender**, v. 6, n. 2, p. 105-114, 2019. https://doi.org/10.1089/vio.2017.0066.