



Trabalho de Conclusão de Curso

**Identificação e Interpretação de Debates  
Feministas no Instagram Brasileiro: Uma  
abordagem com BERTopic e LLaMA**

Thalia Barbosa Marques de Almeida  
tbma@ic.ufal.br

Orientador:  
Prof. Dr. André Luiz Lins de Aquino

Maceió, Novembro de 2024

Thalia Barbosa Marques de Almeida

# **Identificação e Interpretação de Debates Feministas no Instagram Brasileiro: Uma abordagem com BERTopic e LLaMA**

Monografia apresentada como requisito parcial para  
obtenção do grau de Bacharel em Ciência da Com-  
putação do Instituto de Computação da Universidade  
Federal de Alagoas.

Orientador:

Prof. Dr. André Luiz Lins de Aquino

Maceió, Novembro de 2024

**Catálogo na Fonte**  
**Universidade Federal de Alagoas**  
**Biblioteca Central**  
**Divisão de Tratamento Técnico**

Bibliotecário: Jone Sidney A. de Oliveira – CRB-4 - 1485

A447i Almeida, Thalia Barbosa Marques de.  
Identificação e Interpretação de Debates Feministas no Instagram Brasileiro: Uma abordagem com BERTopic e LLaMA / Thalia Barbosa Marques de Almeida. – 2024.  
37 f. : il.

Orientador: André Luiz Lins de Aquino.  
Monografia (Trabalho de conclusão de curso em Engenharia de Computação) - Universidade Federal de Alagoas, Instituto de Computação. Maceió, 2024.  
Texto em inglês.

Bibliografia: f. 33-37.

1. Redes Sociais. 2. Feminismo. 3. Processamento de Linguagem Natural (PLN). 4. BERTopic – LLaMA. 5. Modelagem de Tópicos. I. Título.

CDU: 004.652:396

Monografia apresentada como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação do Instituto de Computação da Universidade Federal de Alagoas, aprovada pela comissão examinadora que abaixo assina.

---

Prof. Dr. André Luiz Lins de Aquino - Orientador  
Universidade Federal de Alagoas

---

Profa. Dra. Fabiane da Silva Queiroz - Examinadora  
Universidade Federal de Alagoas

---

Profa. Dra. Sheyla Christine Santos Fernandes - Examinadora  
Universidade Federal de Alagoas

# Agradecimentos

Agradeço primeiramente a Deus, depois à minha mãe, a quem dedico todas as vitórias da minha vida. Ela não apenas me concedeu a vida, mas também todas as oportunidades de estudo e crescimento que eu poderia ter. Sua dedicação incansável, apoio incondicional e amor foram fundamentais em cada etapa do meu caminho. Às minhas irmãs, Isabella e Thainá, que são minhas melhores amigas e apoiadoras. Sempre me incentivaram a seguir o que amo e torceram por cada conquista, grandes e pequenas. Agradeço profundamente pelo apoio de vocês, certeza de que, em qualquer circunstância, vocês estão ao meu lado. Ao meu pai, pela força e admiração constantes, por todos os gestos e boas comidas que de alguma forma, me fizeram sentir especial.

Aos amigos que cruzaram meu caminho ao longo da vida acadêmica, em especial ao meu amigo de faculdade e da vida, Murilo. Desde que me aceitou em seu grupo para competição de programação, tornou-se minha dupla inseparável em trabalhos e perrengues acadêmicos. Fez dos meus anos na faculdade uma experiência leve e repleta de boas memórias. Às minhas queridas amigas Hellena e Ester, que foram meu suporte feminino em um ambiente predominantemente masculino. Agradeço por todos os trabalhos e experiências compartilhadas, por todos os conselhos, por estarem ao meu lado em cada fase dessa jornada. Sem vocês, a minha graduação beirava o impossível.

Aproveito para agradecer a todo o grupo de extensão Katie e a todas as integrantes que fazem e fizeram este grupo acontecer. Participar desse projeto foi transformador para mim. A sororidade, o altruísmo e a entrega de todas as meninas do Katie são admiráveis, e sinto-me imensamente grata por ter feito parte de algo tão especial.

Finalmente, ao laboratório Orion, especialmente ao coordenador e meu orientador, Professor André Aquino. Sua confiança em mim desde o início, seu cuidado e sua orientação constante foram essenciais para o meu desenvolvimento profissional e acadêmico. Minha gratidão também aos integrantes do laboratório, que tornaram a convivência diária leve e extrovertida, em especial agradeço: Elias, Douglas, Danilo Fernandes, Keila, Rebeca Coli e Rebeca Brandão. Obrigada por todos os momentos e por fazerem dessa experiência uma fase memorável da minha vida.

Thalia Almeida

# Resumo

Este trabalho investiga a identificação e interpretação de temas feministas abordados por perfis brasileiros no Instagram, utilizando técnicas de Processamento de Linguagem Natural (PLN) e modelagem de tópicos. Com o objetivo de analisar discussões recorrentes e suas variações, o estudo aplicou a técnica BERTopic para modelagem de tópicos e o modelo de linguagem de grande escala (LLM) LLaMA para enriquecer a interpretação dos temas identificados. A coleta de dados abrangeu mais de 11 mil postagens de seis perfis feministas, e o pré-processamento foi realizado em Python, utilizando a biblioteca spaCy para tokenização, lematização e remoção de stopwords. A modelagem resultou em 90 tópicos, entre os quais se destacaram temas como violência doméstica, direitos reprodutivos, saúde mental e autonomia do corpo, refletindo debates atuais e relevantes no contexto brasileiro. Para validar a interpretação dos tópicos, um experimento comparativo foi conduzido, avaliando a consistência entre as respostas humanas e as interpretações geradas pelo LLaMA. Embora as métricas de similaridade lexical apresentassem baixa correspondência, o BERTScore indicou uma alta similaridade semântica, sugerindo que o LLaMA capturou o significado dos temas com precisão, confirmando a hipótese alternativa de que técnicas de PLN e LLMs podem identificar e interpretar temas feministas no Instagram de forma eficaz. A relevância desta pesquisa reside em sua contribuição para o campo da análise de discussões sociais em redes digitais, oferecendo uma abordagem estruturada e inovadora para entender debates feministas em uma plataforma de grande alcance como o Instagram. Os resultados reforçam a importância de técnicas de PLN e de modelos de linguagem de grande escala na interpretação de temas sociais complexos, proporcionando uma base sólida para estudos futuros sobre o impacto das redes sociais na conscientização e promoção de mudanças sociais.

**Palavras-chave:** Feminismo, Modelagem de Tópicos, BERTopic, LLaMA, Modelos de Linguagem de Grande Escala (LLMs), Instagram, Processamento de Linguagem Natural (PLN), Redes Sociais

# Abstract

This study investigates the identification and interpretation of feminist themes addressed by Brazilian profiles on Instagram, using Natural Language Processing (NLP) techniques and topic modeling. Aiming to analyze recurring discussions and their variations, the study applied the BERTopic technique for topic modeling and the large language model (LLM) LLaMA to enrich the interpretation of identified themes. Data collection covered over 11,000 posts from six feminist profiles, and preprocessing was conducted in Python using the spaCy library for tokenization, lemmatization, and stopword removal. The modeling identified 90 topics, including prominent themes such as domestic violence, reproductive rights, mental health, and bodily autonomy, reflecting current and relevant debates within the Brazilian context. To validate topic interpretation, a comparative experiment was conducted, evaluating the consistency between human responses and interpretations generated by LLaMA. Although lexical similarity metrics showed low correspondence, the BERTScore indicated high semantic similarity, suggesting that LLaMA effectively captured the meaning of the topics, thus confirming the alternative hypothesis that NLP techniques and LLMs can accurately identify and interpret feminist themes on Instagram. The relevance of this research lies in its contribution to the field of social discourse analysis on digital networks, offering a structured and innovative approach to understanding feminist debates on a widely accessible platform like Instagram. The results underscore the importance of NLP techniques and large language models in interpreting complex social themes, providing a solid foundation for future studies on the impact of social media on awareness and the promotion of social change.

**Key-words:** Feminism, Topic Modeling, BERTopic, LLaMA, Large Language Models (LLMs), Instagram, Natural Language Processing (NLP), Social Media

# Lista de Figuras

4.1	Metodologia aplicada para a modelagem de tópicos abordados em perfis feministas no Instagram. <b>Fonte:</b> Thalia Barbosa Marques de Almeida. Acervo Pessoal. . . . .	14
5.1	Topic Word Scores dos 12 principais tópicos. <b>Fonte:</b> Thalia Barbosa Marques de Almeida. Acervo Pessoal. . . . .	25
5.2	Nuvem de palavras com as principais palavras associadas ao Tópico -1. <b>Fonte:</b> Thalia Barbosa Marques de Almeida. Acervo Pessoal. . . . .	26
5.3	Médias e variabilidade das métricas de coerência e similaridade por tópico. As barras representam as médias das métricas BLEU, ROUGE, Cosine Similarity e BERTScore para cada tema, com barras de erro indicando o desvio padrão, facilitando a comparação entre os diferentes tópicos. <b>Fonte:</b> Thalia Barbosa Marques de Almeida. Acervo Pessoal. . . . .	28
5.4	Comparação de similaridade semântica por tópico (Similaridade de Cosseno vs BERTScore). <b>Fonte:</b> Thalia Barbosa Marques de Almeida. Acervo Pessoal. . .	30
5.5	Médias gerais das métricas com barras de erro que representam o desvio padrão. <b>Fonte:</b> Thalia Barbosa Marques de Almeida. Acervo Pessoal. . . . .	31

# Conteúdo

<b>Lista de Figuras</b>	<b>iii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Contexto . . . . .	1
1.2 Problema de Pesquisa . . . . .	2
1.3 Hipótese . . . . .	3
1.4 Objetivos . . . . .	3
1.4.1 Objetivo Geral . . . . .	3
1.4.2 Objetivos Específicos . . . . .	3
<b>2 Fundamentação Teórica</b>	<b>4</b>
2.1 Processamento de Linguagem Natural (PLN) . . . . .	4
2.1.1 Representação de Texto e Técnicas Tradicionais . . . . .	4
2.1.2 Transformadores . . . . .	5
2.2 Large Language Models (LLMs) . . . . .	5
2.2.1 BERT (Bidirectional Encoder Representations from Transformers) . . . . .	5
2.2.2 LLaMA (Large Language Model Meta AI) . . . . .	6
2.3 Modelagem de Tópicos . . . . .	7
2.3.1 LDA (Latent Dirichlet Allocation) . . . . .	7
2.3.2 Modelagem de Tópicos com Embeddings . . . . .	8
<b>3 Trabalhos Relacionados</b>	<b>9</b>
<b>4 Metodologia</b>	<b>13</b>
4.1 Seleção dos Perfis no Instagram . . . . .	13
4.2 Pipeline de Processamento e Análise de Tópicos usando BERTopic . . . . .	14
4.2.1 Coleta de dados . . . . .	15
4.2.2 Pré-processamento dos Dados . . . . .	15
4.2.3 Modelagem de Tópicos . . . . .	17
4.3 Interpretação dos tópicos . . . . .	19
4.3.1 Experimento de Interpretação Humana via Google Forms . . . . .	19
4.3.2 Interpretação Automatizada com LLaMA . . . . .	20
4.4 Métricas de Avaliação . . . . .	21
<b>5 Resultados e Discussões</b>	<b>24</b>
5.1 Dados Coletados e Preparação para Análise . . . . .	24
5.2 Resultados da Modelagem de Tópicos com BERTopic . . . . .	24
5.2.1 Análise do Tópico -1: Postagens Atípicas ou Outliers . . . . .	25
5.3 Comparação das Interpretações Humanas e do LLM LLaMA . . . . .	26

---

5.3.1	Similaridade entre as respostas humanas com o LLaMA . . . . .	29
<b>6</b>	<b>Conclusão</b>	<b>32</b>
6.1	Trabalhos futuros . . . . .	33
	<b>Referências bibliográficas</b>	<b>34</b>

# 1

## Introdução

### 1.1 Contexto

O feminismo emergiu entre o final do século XVIII e o início do XIX, inicialmente centrado em direitos civis básicos, como o voto e a educação com o movimento sufragista (DuBois, 1975), caracterizando a primeira onda do movimento. Com a conquista desses direitos, o foco se expandiu para incluir a luta contra opressões sociais e econômicas. Na década de 1960, a segunda onda trouxe discussões sobre sexualidade, direitos reprodutivos e violência contra a mulher, criticando o patriarcado como uma estrutura que perpetua a opressão de gênero. A terceira onda, a partir dos anos 1990, introduziu a interseccionalidade, abordando questões de raça e classe, proporcionando uma compreensão mais ampla das diferentes experiências e opressões que afetam mulheres diversas, incluindo as mulheres negras e de classes menos favorecidas.

No Brasil, o autor Diógenes et al. (2015) faz uma reflexão sobre a longa trajetória de luta do movimento feminista brasileiro em busca dos direitos das mulheres. Ao longo dessa caminhada, é possível celebrar várias conquistas importantes, como o avanço das leis e as mudanças nos costumes sociais. No entanto, ao analisar a situação atual, o autor observa que ainda persistem problemas denunciados pelas mulheres do movimento, que continuam a ser uma realidade no Brasil. Esses problemas incluem a violência, tanto no âmbito privado quanto no público, a desigualdade salarial (mesmo com as mulheres apresentando maior qualificação), as múltiplas jornadas de trabalho, que incluem o trabalho doméstico, e a necessidade de resolver questões relacionadas aos direitos reprodutivos e ao respeito à diversidade sexual.

Com tantas causas a serem enfrentadas, essas mulheres precisaram se reinventar a cada revolução para continuar com sua luta e disseminar seus ideais. Na Era da Informação, surgiu um novo conceito denominado Ciberfeminismo (Paasonen, 2011), uma ramificação mais recente do movimento feminista, que explora as possibilidades da internet como ferramenta para

empoderamento e mobilização. Ele surge da fusão entre o feminismo e a tecnologia, questionando a presença feminina no espaço digital e utilizando a rede para discutir questões de gênero, promover ações ativistas e democratizar o acesso ao conhecimento.

O Instagram, com seu caráter altamente visual e sua base global de mais de 1 bilhão de usuários ativos, segundo [Statista \(2024\)](#), tem sido uma plataforma estratégica para a articulação de questões feministas, criando um ambiente no qual esses movimentos podem florescer e alcançar novas audiências. [Mendes et al. \(2018\)](#) destaca como o “*hashtag feminism*” (feminismo de hashtag) tem permitido que mulheres de todo o mundo compartilhem suas experiências e criem uma comunidade virtual de solidariedade. Campanhas virais como #MeToo e #TimesUp, que ganharam imensa visibilidade global, são exemplos disso. ([Saavedra, 2019](#)).

Segundo os estudos de [Banet-Weiser \(2018\)](#), essas plataformas oferecem às mulheres a oportunidade de articular suas vozes de maneira poderosa e visível, sem a necessidade de passar pelos *gatekeepers* ([White, 1950](#)) tradicionais da mídia e da política. [Vachhani \(2024\)](#) enfatiza a importância de compreender as transformações que ocorrem no ambiente digital, tanto em termos de temáticas quanto de políticas, para avaliar o impacto do feminismo digital e suas implicações para as lutas feministas atuais e futuras. Ela destaca que, embora o ativismo feminista tenha historicamente desafiado estruturas de opressão, o ativismo digital introduz novas tensões e contradições que precisam ser compreendidas para construir solidariedade e promover a agência coletiva.

Nos últimos anos, várias pautas referentes ao direito das mulheres foram abordadas no Brasil. Em setembro de 2023, uma importante votação no [Supremo Tribunal Federal \(2023\)](#) se aproximou, com a ministra Rosa Weber votando pela descriminalização do aborto até 12 semanas. Em outro caso, dados atualizados de 2024 do Laboratório de Estudos de Femicídio [LESFEM - Monitor de Femicídios do Brasil \(2024\)](#) da Universidade Estadual de Londrina (UEL) revelou um aumento alarmante nos casos de feminicídios em todo o país, o monitor informa que foram registrados 750 feminicídios consumados e 1693 casos de feminicídios consumados e tentados até junho do mesmo ano. Isso representa uma média diária ainda maior de feminicídio (6,05), em comparação aos casos de 2023 (2,7).

Diante desses fatos, é evidente a necessidade de compreender como os porta-vozes feministas têm respondido a essas questões e de que maneira estão disseminando essas informações para o seu público-alvo na internet.

## 1.2 Problema de Pesquisa

O aumento do uso de redes sociais como o Instagram tem criado uma plataforma significativa para a disseminação de diversos temas sociais e políticos. Entre esses temas, destacam-se as discussões feministas, que, conforme apresentado na Seção 1.1, utilizam a rede para promover debates, conscientizar a população e impulsionar mudanças sociais. No entanto, devido

ao grande volume e à diversidade de conteúdo, identificar e interpretar esses temas de forma estruturada e significativa apresenta-se como um desafio. Diante disso, o problema de pesquisa formulado é: **Como identificar e interpretar os temas abordados por perfis feministas brasileiros no Instagram utilizando técnicas de análise de dados e modelagem de tópicos?**

## 1.3 Hipótese

Para investigar o problema de pesquisa, foi formulada uma hipótese principal que orienta o estudo:

- **Hipótese Nula ( $H_0$ ):** O uso de técnicas de análise de dados e modelagem de tópicos, como o BERTopic, não é suficiente para identificar e interpretar os temas feministas no Instagram, sendo necessária a intervenção humana para uma interpretação precisa dos temas.
- **Hipótese Alternativa ( $H_1$ ):** O uso de técnicas de análise de dados e modelagem de tópicos, como o BERTopic, em combinação com modelos de linguagem de grande escala (LLMs), é suficiente para identificar e interpretar os temas feministas no Instagram, dispensando a necessidade de intervenção humana na interpretação dos temas.

## 1.4 Objetivos

### 1.4.1 Objetivo Geral

Analisar os temas abordados por perfis feministas brasileiros no Instagram, utilizando técnicas de Processamento de Linguagem Natural (PLN) e modelagem de tópicos para identificar e interpretar discussões recorrentes e suas variações.

### 1.4.2 Objetivos Específicos

1. Coletar e organizar dados textuais de postagens e legendas de perfis feministas no Instagram, criando um corpus que permita a análise dos temas.
2. Aplicar métodos tradicionais de modelagem de tópicos, utilizando o modelo BERTopic, para identificar os principais temas abordados pelos perfis.
3. Integrar um modelo de Linguagem de Grande Escala (LLMs), utilizando o modelo de código aberto LLaMA da Meta AI, para enriquecer a análise e obter uma compreensão mais aprofundada dos temas.
4. Comparar os resultados da forma tradicional de interpretação dos tópicos com os obtidos utilizando um LLM, avaliando as vantagens e limitações de cada abordagem.

# 2

## Fundamentação Teórica

Este capítulo apresenta a base conceitual que sustenta o desenvolvimento deste trabalho, oferecendo uma visão ampla e estruturada dos principais conceitos, teorias e estudos que orientam a pesquisa.

### 2.1 Processamento de Linguagem Natural (PLN)

O Processamento de Linguagem Natural (PLN) é um campo de pesquisa que explora métodos para permitir que as máquinas compreendam e manipulem a linguagem humana (Manning and Schutze, 1999). As abordagens de PLN se desenvolveram de técnicas estatísticas e baseadas em regras para métodos mais avançados, impulsionados por redes neurais e grandes conjuntos de dados (Young et al., 2018).

#### 2.1.1 Representação de Texto e Técnicas Tradicionais

##### Bag-of-Words (BoW)

O modelo Bag-of-Words (BoW) é uma das abordagens iniciais mais usadas para representação de texto. BoW representa o texto com base nas palavras presentes, ignorando a ordem das mesmas e focando apenas em suas frequências. Isso gera uma matriz esparsa, onde cada coluna representa uma palavra única do vocabulário, e cada linha corresponde à frequência de ocorrência daquela palavra em cada documento (Harris, 1954). Essa técnica é eficiente para capturar a presença de palavras, porém perde o contexto e sofre com a alta dimensionalidade em vocabulários extensos (Chowdhury, 2010).

## **N-gramas**

Para capturar sequências de palavras, introduzimos o conceito de n-gramas, que representa o texto em blocos consecutivos de “n” palavras (Brown et al., 1992). N-gramas permitem que o modelo capture parte do contexto nas frases, representando melhor a estrutura linguística. Entretanto, esse método aumenta a dimensionalidade e complexidade do modelo conforme o valor de “n” cresce (Parsing, 2009).

## **Embeddings: Representação Distribuída de Palavras**

Modelos de embeddings revolucionaram o PLN ao oferecer representações distribuídas e densas para palavras, que preservam o significado semântico e reduzem a dimensionalidade Mikolov (2013). Em vez de representar palavras como índices em uma matriz esparsa, os embeddings mapeiam palavras para um espaço vetorial de baixa dimensão, no qual palavras semanticamente semelhantes são representadas por vetores próximos.

### **2.1.2 Transformadores**

O modelo de transformadores, introduzido por Vaswani et al. (2017), trouxe avanços significativos no Processamento de Linguagem Natural (PLN) ao utilizar mecanismos de atenção. Esse mecanismo permite que o modelo foque em diferentes partes da entrada, calculando pesos para cada palavra no contexto, capturando assim relações semânticas e sintáticas complexas, inclusive em frases longas. A arquitetura do transformador é composta por um codificador e um decodificador: o codificador gera representações da sentença de origem, e o decodificador as utiliza para produzir a saída, como em tarefas de tradução.

## **2.2 Large Language Models (LLMs)**

Os Large Language Models (LLMs), como BERT (Bidirectional Encoder Representations from Transformers), GPT (Generative Pre-trained Transformer) e LLaMA (Large Language Model Meta AI), são modelos baseados na arquitetura de transformadores e treinados em grandes volumes de dados (Devlin, 2018; Radford, 2018). Esses modelos têm uma grande capacidade de compreensão e geração de linguagem natural, desempenhando uma variedade de tarefas de PLN, como resumo automático, tradução, resposta a perguntas, entre outras.

### **2.2.1 BERT (Bidirectional Encoder Representations from Transformers)**

O BERT, desenvolvido por Devlin (2018), é um modelo bidirecional que considera o contexto de palavras à esquerda e à direita, o que possibilita uma compreensão mais profunda de textos complexos. BERT é amplamente utilizado em tarefas de classificação e entendimento de texto

devido à sua capacidade de capturar contexto bidirecional, ao contrário de modelos unidimensionais.

## 2.2.2 LLaMA (Large Language Model Meta AI)

Desenvolvido pela Meta AI, o LLaMA (Large Language Model Meta AI) é uma série de modelos LLM lançada com o objetivo de democratizar o acesso a LLMs de alto desempenho. Em vez de se concentrar exclusivamente em um modelo massivo, o LLaMA é oferecido em uma gama de tamanhos (de 7 a 65 bilhões de parâmetros), tornando-o acessível para pesquisadores e desenvolvedores com restrições de recursos computacionais. Segundo [Touvron et al. \(2023\)](#), o LLaMA se destaca por sua eficiência, oferecendo desempenho comparável ao de outros grandes modelos, mas com uma arquitetura mais leve que permite seu uso em hardware mais acessível. Essa série de modelos oferece benefícios práticos, especialmente em pesquisas e aplicações onde a capacidade de adaptar o modelo ao tamanho desejado é uma vantagem significativa.

A seguir descrevemos de forma conceitual e algorítmica o funcionamento do LLaMA, incluindo as etapas de tokenização, geração de embeddings, atenção e produção do texto final.

1. **Tokenização do Texto de Entrada:** Cada frase ou parágrafo de entrada é convertido em tokens (fragmentos de texto, como palavras ou sub-palavras) usando uma função de tokenização  $T(x)$ , onde:

$$\text{tokens} = T(x)$$

Aqui,  $x$  representa o texto de entrada.

2. **Geração de Embeddings:** Cada token  $t_i$  é transformado em um vetor de embedding  $e_i$  em um espaço de alta dimensionalidade, usando uma função de embedding  $E(t_i)$ :

$$e_i = E(t_i)$$

Esses embeddings representam o significado semântico dos tokens.

3. **Passagem pelas Camadas de Atenção:** LLaMA aplica múltiplas camadas de autoatenção, que permitem ao modelo focar em diferentes partes do texto conforme a importância de cada token em relação aos outros tokens. O cálculo da atenção para cada token  $e_i$  é dado pela fórmula de atenção escalar:

$$\text{Attention}(Q, K, V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V$$

onde:

- $Q$ ,  $K$ , e  $V$  são as matrizes de consulta, chave e valor derivadas dos embeddings,
- $d_k$  é a dimensão de  $K$ , usada para normalização.

4. **Composição e Geração do Texto:** Após passar pelas camadas de atenção, o modelo gera uma probabilidade  $p(y|x)$  para cada possível próximo token  $y$  dado o contexto  $x$ . O token mais provável é selecionado em cada etapa do processo de geração, usando:

$$y_i = \arg \max p(y|x_{1:i-1})$$

Este processo continua iterativamente até que uma condição de parada seja atingida (por exemplo, quando um token de final de frase é gerado).

5. **Produção do Texto Final:** Finalmente, o texto gerado é o resultado da concatenação dos tokens  $y_i$ , formando uma frase ou parágrafo.

Esse processo permite que o LLaMA capture contextos complexos e forneça respostas coerentes, aplicando técnicas avançadas de aprendizado de máquina e atenção contextual para interpretar e gerar linguagem natural com precisão.

O processo de geração de texto com o LLaMA pode ser resumido no seguinte algoritmo:

---

**Algoritmo 2.1** Algoritmo para Geração de texto com LLaMA

---

**Input:**  $x$  (texto de entrada)

**Output:** Geração de texto pelo LLaMA

- 1 Aplicar tokenização:  $\text{tokens} = T(x)$  Calcular embeddings:  $e_i = E(t_i)$  para cada token  $t_i$  Passar embeddings pelas camadas de atenção para contextualização semântica **for** cada token de saída  $y_i$  **do**
  - 2 | Gerar  $y_i = \arg \max p(y|x_{1:i-1})$  usando o modelo autoregressivo
  - 3 **end**
  - 4 Concatenar  $y_i$  para produzir o texto final
- 

## 2.3 Modelagem de Tópicos

A modelagem de tópicos é uma técnica usada para descobrir temas latentes em grandes coleções de textos, agrupando palavras que frequentemente aparecem juntas para identificar temas principais (Blei et al., 2003). Com o desenvolvimento dos LLMs, novos métodos de modelagem de tópicos foram criados para aproveitar as representações contextuais ricas desses modelos.

### 2.3.1 LDA (Latent Dirichlet Allocation)

Um dos métodos mais comuns para modelagem de tópicos é o LDA (Latent Dirichlet Allocation), um modelo probabilístico que identifica tópicos com base na coocorrência de palavras em documentos (Blei et al., 2003). No entanto, o LDA possui limitações, como a dificuldade em capturar relações complexas entre palavras e a falta de sensibilidade ao contexto (Boyd-Graber et al., 2017).

### **2.3.2 Modelagem de Tópicos com Embeddings**

Com os embeddings, a modelagem de tópicos evoluiu para técnicas que combinam embeddings contextuais com algoritmos de clustering, permitindo uma identificação de tópicos mais precisa e sensível ao contexto (Sia et al., 2020). Que são posteriormente agrupadas por métodos como K-means ou DBSCAN para identificar grupos de tópicos com base em semelhanças semânticas.

# 3

## Trabalhos Relacionados

Este capítulo explora os principais estudos relacionados à modelagem de tópicos, analisando os métodos mais utilizados, suas limitações e os avanços recentes que propõem novas abordagens.

Estudos sobre a modelagem de tópicos em textos curtos extraídos de redes sociais como o de [Muthusami et al. \(2024\)](#), evidenciam os desafios de capturar a riqueza e a complexidade das narrativas presentes nesses dados.

A natureza fragmentada e concisa dos textos em plataformas como Instagram e Twitter, onde postagens e legendas são frequentemente limitadas em comprimento e acompanhadas de imagens e *hashtags*, impõe obstáculos significativos para métodos tradicionais de modelagem de tópicos, como o *Latent Dirichlet Allocation* (LDA) ([Blei et al., 2003](#)), que foi projetado para analisar corpus de textos mais longos e coerentes. O estudo de [Mazarura and de Waal \(2016\)](#) compara dois modelos de tópicos aplicados a textos curtos: o LDA e o *Dirichlet Multinomial Mixture Model* (GSDMM). Ele destaca que, embora o LDA funcione bem para textos longos, como livros e artigos, seu desempenho é inferior em textos curtos, como postagens em redes sociais. O GSDMM, por outro lado, parece ser mais adequado para textos curtos, pois assume que cada documento pertence a apenas um tópico.

Além disso, a variabilidade da linguagem usada nas redes sociais, incluindo gírias, emojis e *hashtags*, acrescenta uma camada adicional de complexidade. Essas características linguísticas e visuais frequentemente não são bem capturadas pelos métodos tradicionais de modelagem de tópicos, que podem não interpretar corretamente a importância ou o significado desses elementos dentro do texto ([Blei and Lafferty, 2007](#)).

Diversas aplicações de extração de tópicos foram identificadas na literatura. Esses estudos buscam fazer inferências sobre ambientes do mundo real, explorando a relação entre tópicos e outros fatores contextuais como em [Joo et al. \(2020\)](#). [Kurten and Beullens \(2021\)](#) que queriam determinar se o número de *tweets* variava conforme o cronograma da pandemia e suas etapas, além de analisar como o conteúdo desses *tweets* evoluía ao longo do tempo. Em outro estudo, [Nobles et al. \(2020\)](#) utilizaram a modelagem de tópicos para compreender como indivíduos que

se autoidentificavam como portadores de Vírus da Imunodeficiência Humana (HIV) expressavam suas experiências com a doença.

O processo de preparação dos dados, especialmente os provenientes de redes sociais, é abordado de diferentes maneiras na literatura. No pré-processamento, é comum a remoção de e-mails, URLs e elementos *HyperText Markup Language* (HTML) dos textos (Bahja and Safdar, 2020; Cai et al., 2020). Na normalização dos dados, diferentes abordagens têm sido adotadas, um exemplo é a remoção de palavras-chave usadas para coletar os dados. Segundo o artigo de revisão Laureate et al. (2023), a remoção dessas palavras é essencial, pois, caso contrário, os tópicos gerados tendem a ser dominados pelos termos de consulta. Por exemplo, Xu and Xiong (2020) analisaram tweets sobre a polêmica campanha de marketing da Gillette em 2019, utilizando a *hashtag* #gillette. No entanto, como não removeram a hashtag nem o termo "Gillette" das análises, é provável que os tópicos identificados tenham sido fortemente influenciados por esses termos, comprometendo a análise.

Na modelagem dos tópicos, diferentes métodos são empregados. De acordo com Laureate et al. (2023), cerca de 79,79% dos artigos analisados até 2021 utilizaram o LDA como modelo de geração de tópicos, seguido pelo Structural Topic Model (STM), com 6,74%. O restante utilizava uma combinação de LDA com outro modelo.

Com o avanço das técnicas de modelagem de tópicos, um estudo publicado em 2022 por Grootendorst (2022) introduziu o BERTopic, um modelo que combina *embeddings* de linguagem, especificamente o *Bidirectional Encoder Representations from Transformers* (BERT) (Kenton and Toutanova, 2019), com técnicas de agrupamento hierárquico e redução dimensional para gerar tópicos mais coerentes. Ao contrário de modelos tradicionais, como o LDA, que utilizam a abordagem de saco de palavras (*bag-of-words*) (Blei et al., 2003), desconsiderando as relações semânticas entre as palavras, a pesquisa afirma que o BERTopic se destaca por capturar o contexto das palavras por meio da geração de *embeddings*, permitindo uma compreensão mais profunda e precisa dos tópicos.

O mesmo estudo Grootendorst (2022) comparou o BERTopic com métodos tradicionais de modelagem de tópicos, como LDA, NMF, CTM e Top2Vec, utilizando três conjuntos de dados: **20 NewsGroups**, **BBC News** e os **tweets de Donald Trump**. As comparações foram realizadas considerando as métricas de coerência e diversidade de tópicos, que, segundo a revisão Laureate et al. (2023), são amplamente usadas para avaliar a qualidade dos modelos. Como observado na Tabela 3.1, o BERTopic obteve altas pontuações em coerência de tópicos em todos os conjuntos de dados, destacando-se especialmente no dataset dos tweets de Trump, que foi ligeiramente pré-processado. Isso sugere que o BERTopic é particularmente eficaz em lidar com dados curtos e menos estruturados, como textos de redes sociais.

No entanto, apesar de seu bom desempenho em coerência, a pesquisa pontua que o BERTopic foi consistentemente superado pelo Correlated Topic Model (CTM) em termos de diversidade de tópicos, um aspecto importante para avaliar a capacidade de um modelo de gerar tópicos variados e não redundantes. O estudo também destaca que essas métricas podem não

	<b>20 NewsGroups</b>		<b>BBC News</b>		<b>Trump</b>	
	TC	TD	TC	TD	TC	TD
LDA	.058	.749	.014	.577	-.011	.502
NMF	.089	.663	.012	.549	.009	.379
T2V- <i>MPNET</i>	.068	.718	-.027	.540	-.213	.698
T2V- <i>Doc2Vec</i>	.192	.823	.171	.792	-.169	.658
CTM	.096	.886	.094	.819	.009	.855
BERTopic- <i>MPNET</i>	.166	.851	.167	.794	.066	.663

Tabela 3.1: Comparação da coerência de tópicos (TC) e diversidade de tópicos (TD) entre diferentes modelos de geração de tópicos, incluindo o BERTopic. A análise foi realizada variando o número de tópicos de 10 a 50, com incrementos de 10. Em cada intervalo, a coerência e a diversidade de tópicos foram calculadas e os resultados foram a média de três execuções. No total, cada pontuação representa a média de 15 execuções separadas.

Fonte: Adaptado de Grootendorst (2022)

representar com precisão a avaliação de um modelo, funcionando apenas como aproximações para algo que é essencialmente subjetivo. Embora o Normalized Pointwise Mutual Information (NPMI) (Bouma, 2009) seja amplamente utilizado como métrica de coerência, alguns estudos como em Hoyle et al. (2021) sugerem que essa métrica é mais adequada para modelos clássicos, como o LDA, e menos precisa para modelos baseados em redes neurais, como o BERTopic.

Além disso, a comparação feita no estudo destaca que o tempo de processamento do BERTopic foi competitivo em relação a outros modelos, o que é um fator importante em aplicações práticas, especialmente quando grandes volumes de dados precisam ser analisados.

Por fim, Grootendorst (2022) observa que, apesar das inovações trazidas pelo BERTopic, uma de suas principais limitações está na representação dos tópicos. Embora o modelo utilize *embeddings* contextuais avançados, como o BERT, para gerar representações semânticas dos documentos, a representação final dos tópicos ainda se baseia em uma abordagem de *bag-of-words*. Isso resulta em listas de palavras que indicam apenas a relevância individual de cada termo no tópico, sem capturar o contexto completo em que essas palavras estão inseridas. Como consequência, as palavras que compõem um tópico podem ser redundantes ou muito semelhantes, o que torna a interpretação dos temas menos precisa. O autor sugere que, embora técnicas como a *Maximal Marginal Relevance* (MMR) pudessem ser aplicadas para diversificar as palavras mais representativas de cada tópico, essa solução não foi explorada no estudo, deixando uma lacuna importante na melhoria da interpretação dos tópicos gerados.

Na etapa de avaliação da qualidade dos tópicos, é comum encontrar na literatura estudos que focam na interpretabilidade dos tópicos gerados, como nos trabalhos de Amin et al. (2020) e Okon et al. (2020). A maioria dos autores baseia essa avaliação na análise dos principais termos associados a cada tópico, como forma de interpretá-los (Hemmatian et al., 2019; Hacker et al., 2020). Nesse processo, os dez termos prováveis são examinados, e um ou mais anotadores atribuem um rótulo ao tópico com base nesses termos conforme descrito por Liang et al. (2019) e Kurten and Beullens (2021). Em muitos estudos, após a rotulação dos tópicos, eles são descritos com base no conhecimento dos autores sobre o conjunto de dados, o assunto ou

outros conhecimentos contextuais como relatado por [Ibrahim and Wang \(2019\)](#).

Esse método é proeminente em vários estudos que aplicam a modelagem de tópicos. Segundo [Laureate et al. \(2023\)](#), cerca de 63,64% das pesquisas analisadas utilizaram essa abordagem de análise. No entanto, vários estudos como em [Brown \(2019\)](#) e [Bérubé et al. \(2020\)](#) expressaram preocupações sobre possíveis vieses e a profundidade dos *insights* obtidos com esse método.

É por esse motivo que [Ibrahim and Wang \(2019\)](#) mencionam que futuras pesquisas devem considerar esses aspectos críticos e explorar métodos adicionais ou alternativos para a interpretação de tópicos.

Pesquisas recentes têm explorado o uso de *Large Language Models* (LLMs), como o GPT-4 e LLaMA-3, na análise de textos. Por exemplo, o estudo [Kwon et al. \(2024\)](#) analisou 1,26 milhão de tweets sobre energia nuclear nos EUA entre 2008 e 2023, utilizando modelos de linguagem para classificar os sentimentos do público sobre temas políticos e energéticos.

Em outro estudo, [Bauer et al. \(2024\)](#) analisou discussões sobre suicídio em *subreddits*, utilizando o modelo de linguagem generativa GPT-4 da OpenAI para identificar padrões latentes na linguagem dos usuários. Os pesquisadores exploraram como sentimentos de desconexão e desesperança são expressos nas postagens, oferecendo *insights* valiosos sobre a comunicação em torno da saúde mental.

São poucos os estudos que exploram o uso de LLMs na modelagem de tópicos. O estudo [Kirilenko and Stepchenkova \(2024\)](#) focou em analisar a capacidade dos grandes modelos de linguagem (LLMs), como o ChatGPT, para extrair os temas principais das reações de espectadores a vídeos populares sobre um destino rural na China. Comparando os resultados com o modelo LDA, os autores verificaram que os LLMs foram mais precisos, específicos e eficazes na separação dos tópicos de discussão, superando as limitações comuns dos métodos tradicionais ao lidar com documentos curtos, como feedbacks de leitores de blogs.

Já [Rijcken et al. \(2023\)](#) explorou uma nova abordagem para interpretar melhor os resultados de modelos de tópicos, utilizando o modelo de linguagem ChatGPT. O estudo demonstrou que o LLM pode oferecer descrições precisas e *insights* valiosos, desde que adequadamente orientado por prompts.



# Metodologia

Este capítulo apresenta em detalhes a metodologia adotada para identificar e interpretar os temas abordados em perfis feministas brasileiros no Instagram. A Figura 4.1 ilustra o processo, estruturado em cinco etapas principais: coleta de dados, pré-processamento dos textos, modelagem de tópicos e interpretação dos tópicos, dividida em duas abordagens: interpretação humana e interpretação realizada por um LLM (Large Language Model).

## 4.1 Seleção dos Perfis no Instagram

A identificação dos perfis de interesse foi realizada por meio dos descritores “feminismo” e “feminista” na barra de busca da plataforma Instagram. Foram considerados apenas perfis que abordassem diretamente o tema do estudo, fossem de acesso público e apresentassem uma quantidade significativa de seguidores, estipulada em pelo menos 10 mil. Os perfis selecionados representam uma variedade de visões e abordagens sobre questões feministas, com atividades recentes nos últimos 12 meses. A lista de perfis escolhidos foram:

- @planetaella
- @feminiismo
- @coletivo\_feminista
- @feminismo\_sendemagogia
- @arquivosfeministas
- @revistatpm

A partir da seleção desses perfis, foi possível coletar dados relevantes para a análise dos discursos feministas presentes na plataforma. Com o objetivo de identificar os principais tópicos abordados, realizou-se um processamento textual dos dados extraídos. Na seção a seguir, descrevemos o pipeline utilizado para o processamento e análise de tópicos, realizado com o auxílio do modelo BERTopic.

## 4.2 Pipeline de Processamento e Análise de Tópicos usando BERTopic

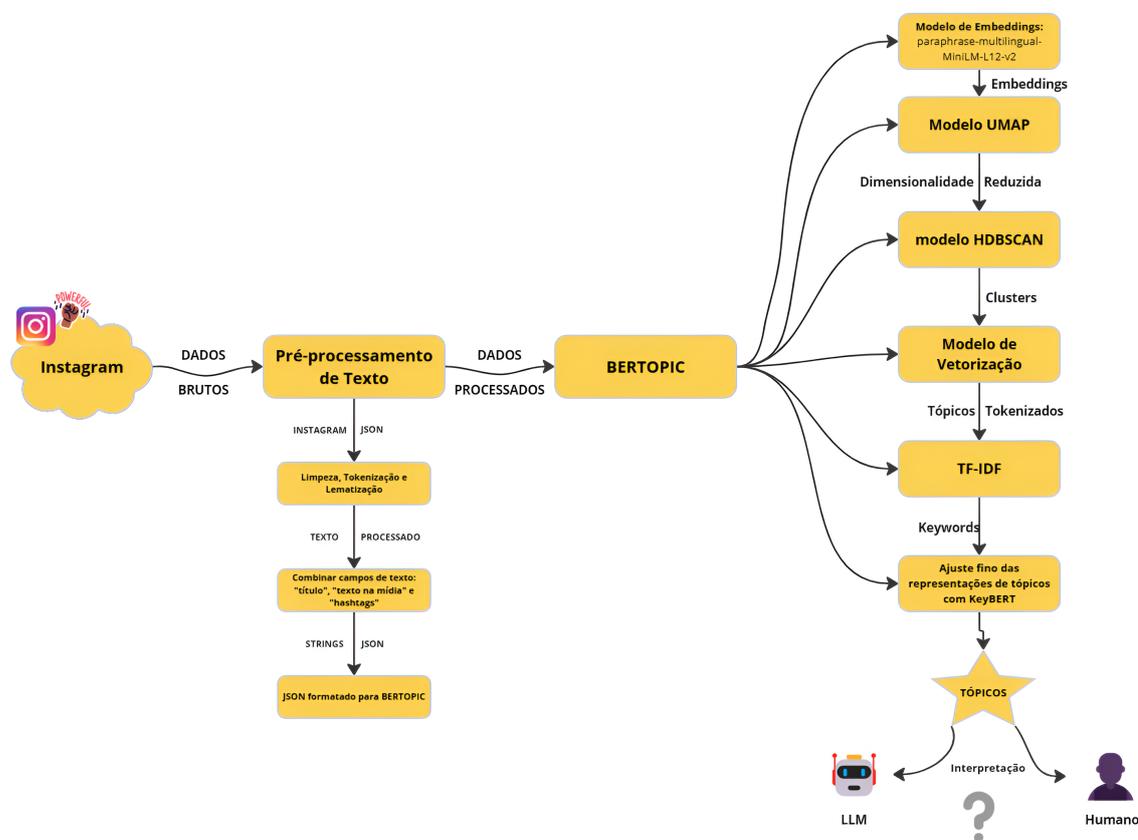


Figura 4.1: Metodologia aplicada para a modelagem de tópicos abordados em perfis feministas no Instagram. **Fonte:** Thalia Barbosa Marques de Almeida. Acervo Pessoal.

A Figura 4.1 ilustra o fluxo de etapas envolvidas neste pipeline, desde a pré-processamento dos dados até a análise e interpretação dos tópicos gerados. Cada etapa é projetada para transformar e organizar os dados, tornando-os adequados para a modelagem de tópicos e facilitando a identificação de padrões e temas recorrentes. Abaixo, cada uma dessas etapas é explicada em detalhes.

### 4.2.1 Coleta de dados

A coleta de dados foi realizada utilizando uma *Application Programming Interface* (API) de *Open Source Intelligence* (OSINT), que faz chamadas à API do Instagram para extrair informações públicas das postagens dos perfis selecionados. Foram coletadas postagens (legendas), *hashtags* (etiqueta ou marcador de palavra-chave, frequentemente usada em redes sociais para categorizar conteúdos), textos contidos nas mídias (como imagens e vídeos), e outros dados públicos disponíveis através da API do Instagram. Todas as postagens coletadas datam de um período de um ano ou mais recente, visando garantir a atualidade e relevância das narrativas feministas no contexto brasileiro.

### 4.2.2 Pré-processamento dos Dados

O pré-processamento dos dados foi implementado por meio de uma classe em Python, utilizando a biblioteca *spaCy* para as operações de tokenização, lematização e remoção de *stopwords*. Os dados extraídos da API foram retornados no formato *JavaScript Object Notation* (JSON) e, para garantir a qualidade e a consistência das informações, o pré-processamento foi dividido em quatro etapas principais.

Cada etapa foi desenvolvida para preparar o conjunto de dados de forma adequada para a análise com o BERTopic.

#### Filtragem das Informações

A primeira etapa do pré-processamento consistiu em filtrar as informações mais relevantes para o objetivo da pesquisa. Embora a API retorne diversos tipos de dados, foram selecionados apenas os seguintes campos para compor a análise:

- **Legenda do post (title):** O texto principal que acompanha as postagens e serve como a descrição central do conteúdo.
- **Textos nas imagens (text\_on\_media):** Algumas postagens incluem informações diretamente nas mídias, como textos contidos a imagem. Esses textos foram extraídos para complementar a análise.
- **Hashtags:** As hashtags associadas a cada post foram também incluídas como elementos adicionais

#### Limpeza dos Textos

Após a filtragem das informações, foi realizada uma limpeza textual para remover elementos irrelevantes e preparar os dados para as etapas subsequentes. A limpeza incluiu as seguintes operações:

- **Decodificação de caracteres:** Os textos recebidos da API vinham com caracteres codificados, principalmente em relação a acentuações da língua portuguesa. Esses caracteres foram decodificados para garantir a legibilidade dos textos.
- **Remoção de emojis:** Um método foi implementado para identificar e remover códigos de emojis, que não contribuía para a análise textual.
- **Remoção de quebras de linha:** Textos brutos incluíam quebras de linha explícitas (`\n`), portanto foram removidas.
- **Remoção de URLs:** URLs, que frequentemente apareciam nas postagens, foram eliminadas, pois não tinham relevância para a análise dos temas.
- **Remoção de menções:** Palavras precedidas de "@" (menções a outros perfis) foram removidas, visto que não contribuía para a compreensão do conteúdo textual.
- **Remoção de pontuação:** Toda a pontuação foi eliminada para garantir um texto mais uniforme.
- **Conversão para minúsculas:** Todos os textos foram convertidos para minúsculas para evitar distinções desnecessárias entre palavras idênticas escritas de forma diferente.

### Tokenização e Normalização com `spaCy`

A próxima etapa foi a *tokenização* e a normalização dos dados, utilizando as bibliotecas `spaCy` e `nltk`. Essa etapa serviu para dividir os textos em unidades linguísticas (tokens) e preparar os dados para a modelagem de tópicos. As operações realizadas incluem:

- **Lematização:** Com o auxílio do `spaCy`, a lematização foi utilizada para reduzir as palavras à sua forma base (ou lema), removendo variações morfológicas (como plural ou conjugação verbal). Isso assegurou que diferentes formas de uma mesma palavra fossem tratadas de forma unificada, evitando duplicidades desnecessárias.
- **Remoção de *Stopwords*:** Com o uso da biblioteca `spaCy`, as *stopwords* (palavras comuns que não agregam valor semântico, como "de", "a", "e") foram removidas, garantindo que os tokens representassem apenas termos relevantes para a modelagem de tópicos.
- **Stemming:** Posteriormente, aplicou-se o *stemming*, usando a biblioteca `nltk`, que simplifica as palavras ao reduzir seus sufixos e variações, transformando-as em suas formas básicas. Essa técnica foi utilizada para padronizar o vocabulário e minimizar a variação de palavras semelhantes, contribuindo para uma maior consistência nos tópicos gerados. Quando combinado com a lematização, o uso do stemming aprimorou significativamente a coerência e a qualidade dos tópicos identificados.

### Formatação da Saída

A última etapa do pré-processamento foi a formatação dos dados para que estivessem prontos para serem utilizados pelo modelo BERTopic. O formato exigido pelo BERTopic é uma lista de *strings*, onde cada *string* representa um post completo. Para isso, foram realizadas as seguintes ações:

- **Combinação das informações:** As diferentes partes de cada postagem (legenda, textos nas imagens e hashtags) foram combinadas em uma única *string* por post.
- **Criação de uma lista de *strings*:** As *strings* de cada post foram unificadas em uma única lista de *strings*.
- **Remoção de *strings* vazias:** Postagens que não continham informações textuais após o pré-processamento foram eliminadas, resultando na remoção de *strings* vazias da lista final.

### 4.2.3 Modelagem de Tópicos

Como destacado nos trabalhos recentes da seção 2.3.2, a abordagem com a técnica de modelagem de tópicos chamada BERTopic é a mais recomendada para análises utilizando textos curtos de mídias sociais, como é o caso das postagens no Instagram. A escolha do BERTopic foi motivada pela sua capacidade de gerar representações semânticas robustas e coerentes mesmo com textos curtos e fragmentados, o que torna essa técnica mais adequada do que métodos tradicionais como LDA.

#### BERTopic

O BERTopic é uma técnica de modelagem de tópicos baseada em transformadores, que utiliza representações de texto geradas por modelos de embeddings combinados com técnicas de redução de dimensionalidade e agrupamento. As etapas principais da modelagem com BERTopic incluem a extração de embeddings, redução de dimensionalidade, agrupamento de dados, tokenização, e a criação de representações de tópicos. As etapas foram configuradas com os valores padrão do BERTopic, com exceção de alguns parâmetros personalizados, como descrito abaixo:

- **Etapa 1 - Embeddings:** Para gerar as representações vetoriais dos textos, foi utilizado o modelo `paraphrase-multilingual-MiniLM-L12-v2`, da biblioteca `SentenceTransformer`, treinado para diversos idiomas, incluindo o português. Esse modelo gera embeddings que capturam o significado semântico dos textos de forma eficiente, facilitando o agrupamento posterior.

- **Etapa 2 - Redução de Dimensionalidade:** Após a extração dos embeddings, a técnica de redução de dimensionalidade UMAP foi aplicada com os parâmetros padrão do BERTopic. Essa técnica permite representar os dados em uma dimensão mais baixa, preservando as relações semânticas.
- **Etapa 3 - Agrupamento:** Para agrupar os textos semanticamente semelhantes, foi utilizado o algoritmo HDBSCAN também com parâmetros padrão do BERTopic. Esse algoritmo identifica automaticamente o número de clusters (tópicos), adaptando-se aos dados sem necessidade de pré-definir a quantidade de tópicos.
- **Etapa 4 - Tokenização dos Tópicos:** A tokenização foi realizada utilizando o `CountVectorizer`, personalizado com uma lista de *stopwords* em português. O BERTopic, por padrão, utiliza *stopwords* em inglês, portanto, a inclusão de *stopwords* específicas da língua portuguesa foi necessária para remover palavras irrelevantes e melhorar a precisão dos tópicos gerados.
- **Etapa 5 - Representação dos Tópicos:** A representação dos tópicos foi realizada utilizando o `ClassTfidfTransformer`, que converte a frequência das palavras em valores ponderados de TF-IDF (Term Frequency - Inverse Document Frequency). Essa técnica é útil para identificar as palavras mais relevantes dentro de cada tópico, garantindo que as palavras-chave representem com precisão o conteúdo dos textos agrupados. O cálculo do TF-IDF é feito conforme a fórmula:

$$\text{TF-IDF}(t, d, D) = \text{TF}(t, d) \times \text{IDF}(t, D)$$

onde:

- $\text{TF}(t, d)$  representa a frequência da palavra  $t$  no documento  $d$ ,
- $\text{IDF}(t, D) = \log\left(\frac{N}{|\{d \in D : t \in d\}|}\right)$ , em que  $N$  é o número total de documentos e  $|\{d \in D : t \in d\}|$  é o número de documentos que contêm a palavra  $t$ .

Essa abordagem pondera as palavras de modo que aquelas mais específicas a cada tópico obtenham maior destaque, o que melhora a representatividade e a precisão na definição de cada tema.

- **Etapa 6 - Ajuste Fino dos Tópicos:** Na etapa final, foi utilizado o método opcional `KeyBERTInspired` para ajustar as palavras-chave dos tópicos. Esse método refina a escolha das palavras-chave com base na similaridade semântica, garantindo que os tópicos sejam descritos por termos mais precisos e representativos.

## 4.3 Interpretação dos tópicos

A modelagem de tópicos com o BERTopic organiza documentos em grupos de acordo com sua similaridade semântica, associando a cada grupo um conjunto de palavras-chave representativas. Entretanto, para interpretar com precisão o significado de cada tópico, é necessário ir além da simples identificação desses grupos e palavras-chave.

A interpretação dos tópicos é uma etapa crucial, pois traduz as palavras-chave geradas automaticamente em temas compreensíveis e aplicáveis ao objetivo da pesquisa. Embora o modelo BERTopic forneça as palavras mais relevantes para cada tópico, estas sozinhas não capturam a complexidade das discussões. Assim, torna-se necessário realizar uma análise interpretativa para atribuir um significado claro e coerente aos tópicos, relacionando-os aos temas abordados nas postagens analisadas.

### 4.3.1 Experimento de Interpretação Humana via Google Forms

Para validar a interpretação dos tópicos, foi realizado um experimento de interpretação com a participação de respondentes. Um formulário foi criado no Google Forms com 10 tópicos selecionados, sendo os 5 primeiros os tópicos mais relevantes (com maior número de postagens) e os 5 restantes representando tópicos diversos, escolhidos aleatoriamente entre os tópicos gerados.

No formulário, foram fornecidas as seguintes informações para cada tópico:

- As 10 palavras-chave mais relevantes que representavam o tópico, calculadas com o método TF-IDF.
- Um exemplo de legenda de postagem associada ao tópico, para ajudar a contextualizar o tema.
- Uma pergunta aberta solicitando ao participante que sugerisse um nome para o tópico, baseado nas palavras-chave e na legenda fornecida.

#### Exemplo da Pergunta Feita aos Respondentes

##### 1. Palavras-chave que representam o tópico:

*Palavras-chave: estupro, aborto, feminismo, feminista, violência, afirmar, gênero, menina, legal, crime.*

**Trecho da legenda de uma postagem:** *"NÃO É UM CASO ISOLADO | Entenda quais os direitos de crianças e adolescentes nessa situação no vídeo acima com anônimo, diretora do Instituto Liberta e doutora especialista em Direito Constitucional. Mais de 21 mil meninas entre 10 e 14 anos engravidam por ano no Brasil..."*

**Pergunta:** Como você nomearia este tópico? (pode ser uma palavra ou uma frase)

### 4.3.2 Interpretação Automatizada com LLaMA

Além da interpretação realizada por participantes humanos, utilizamos o modelo de linguagem de grande escala LLaMA, desenvolvido pela Meta AI, para conduzir uma interpretação automatizada dos tópicos. O objetivo foi avaliar a capacidade do LLaMA de interpretar e sugerir nomes para os tópicos de maneira semelhante à realizada pelos participantes humanos.

#### Criação do Prompt

Para a interpretação automatizada, foi necessário enviar um prompt estruturado ao modelo LLaMA, especificando claramente as instruções e o formato esperado de resposta.

O exemplo abaixo ilustra o formato do *prompt* utilizado:

#### Exemplo de Prompt Enviado ao LLaMA

Você é um assistente que identifica tópicos com base em uma lista de palavras-chave e um texto representativo de tokens fornecido.

Eu vou fornecer uma lista de palavras-chave e uma string de tokens representando o contexto do tópico. Sua tarefa é analisar a lista de palavras-chave e o texto representativo para identificar o tópico mais relevante.

1. Analise a lista de palavras-chave e a string de tokens associada.
2. Use as palavras-chave e o texto representativo para identificar o tópico mais relevante.
3. Retorne um JSON que associe o índice ao nome do tópico correspondente.

#### Modelo de Entrada:

```
{
  "palavras-chave": ["palavra1", "palavra2"],
  "texto": "string de tokens"
}
```

#### Formato de Resposta:

```
{
  "0": "Nome do tópico"
}
```

Agora, analise esta entrada:

O exemplo de prompt utilizado aplicou o conceito de *prompt engineering*, utilizando o for-

mato de *chat* com dois componentes principais: **role** e **content**. O **role** define o papel do LLM no diálogo (neste caso, como um assistente ou especialista), enquanto o **content** fornece as informações necessárias para o modelo realizar a interpretação (White et al., 2023). O modelo foi configurado em um formato de *few-shot learning*, apresentando exemplos no próprio prompt para ajudar o LLaMA a entender a tarefa.

As entradas enviadas ao LLaMA foram fornecidas no formato JSON, derivadas diretamente da saída do BERTopic, facilitando a integração entre os modelos.

Abaixo está um trecho de código que demonstra como cada item (representações de um tópico) foi concatenado ao *prompt*:

#### Exemplo de Concatenamento do Prompt

```
prompt = PROMPT_IDENTIFICACAO_TOPICO + json.dumps(item)
```

## 4.4 Métricas de Avaliação

Para avaliar os resultados, foram utilizadas métricas objetivas, atendendo à necessidade crescente de avaliações automáticas que dispensem o julgamento humano, conforme discutido no artigo de Saeki et al. (2024). Durante muito tempo, avaliações subjetivas foram consideradas o padrão-ouro na avaliação de geração de fala e outros modelos de linguagem; no entanto, o uso de métricas objetivas, como o BERTScore, BLEU e ROUGE, tem se mostrado vantajoso em termos de eficiência de custo e tempo, além de permitir uma avaliação mais consistente em larga escala. Abaixo, detalhamos as ferramentas e configurações utilizadas para cada métrica.

- **BLEU (Bilingual Evaluation Understudy)**: O cálculo do score BLEU foi realizado com a função `sentence_bleu` da biblioteca `nlk.translate`. Esta métrica avalia a precisão entre n-gramas das interpretações, oferecendo uma visão sobre a similaridade léxica exata entre as frases geradas por humanos e pelo LLaMA. A métrica BLEU é amplamente utilizada na tradução automática, pois permite avaliar o quanto o texto gerado coincide com o texto de referência em termos de n-gramas, destacando a similaridade lexical (Papineni et al., 2002).

$$\text{BLEU} = \text{BP} \times \exp \left( \sum_{n=1}^N w_n \log p_n \right)$$

onde:

- BP é o fator de penalização de brevidade, que é igual a 1 se o comprimento da sequência gerada for maior ou igual ao da sequência de referência, e  $\exp(1 - \frac{\text{ref}}{\text{cand}})$  caso contrário.

- $p_n$  representa a precisão dos n-gramas de ordem  $n$ .
- $w_n$  são os pesos para cada ordem de n-gramas, geralmente iguais para um BLEU de 4-gramas.

• **ROUGE (Recall-Oriented Understudy for Gisting Evaluation):** Para o cálculo do ROUGE, foi utilizado o `rouge_scorer` da biblioteca `rouge_score`. Essa função permite medir a sobreposição de palavras entre as interpretações, avaliando a recuperação de termos importantes entre os textos. A métrica ROUGE, desenvolvida por Lin (2004), é frequentemente empregada na avaliação de resumos automáticos, pois mede a capacidade do texto gerado de recuperar informações essenciais do texto de referência, o que é particularmente útil para entender a relevância lexical.

- **ROUGE-N:** Mede a sobreposição de n-gramas entre o texto de referência e o texto gerado, conforme a fórmula:

$$\text{ROUGE-N} = \frac{\sum_{S \in \text{Referência}} \sum_{\text{grama}_n \in S} \text{Contagem}_{\text{match}}(\text{grama}_n)}{\sum_{S \in \text{Referência}} \sum_{\text{grama}_n \in S} \text{Contagem}(\text{grama}_n)}$$

onde  $\text{Contagem}_{\text{match}}(\text{grama}_n)$  é o número de n-gramas em comum entre o texto de referência e o texto gerado.

- **ROUGE-L:** Calcula a similaridade com base na maior subsequência comum (LCS - Longest Common Subsequence) entre o texto de referência e o texto gerado, de acordo com a fórmula:

$$\text{ROUGE-L} = \frac{\text{LCS}(\text{Referência}, \text{Candidato})}{\text{Comprimento da Referência}}$$

onde  $\text{LCS}(\text{Referência}, \text{Candidato})$  representa o comprimento da maior subsequência comum entre o texto de referência e o texto gerado.

• **Cosine Similarity:** A similaridade do cosseno foi calculada com o uso das classes `TfidfVectorizer` e `cosine_similarity` da biblioteca `sklearn`. Inicialmente, as interpretações foram transformadas em vetores TF-IDF, que foram então comparados para medir o ângulo entre os vetores de cada interpretação, indicando a proximidade semântica. Conforme discutido por Rahutomo et al. (2012), a Similaridade do Cosseno é uma métrica eficaz para avaliar a similaridade semântica em espaço vetorial, sendo amplamente utilizada na comparação de documentos.

$$\text{Cosine Similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

onde:

- $A \cdot B$  é o produto escalar dos vetores  $A$  e  $B$ ,
  - $\|A\|$  e  $\|B\|$  são as normas (ou magnitudes) dos vetores  $A$  e  $B$ ,
  - $\theta$  é o ângulo entre os vetores.
- **BERTScore:** Para calcular o BERTScore, foi utilizada a função `score` da biblioteca `bert_score`. Essa métrica calcula a similaridade semântica profunda entre os textos usando embeddings do modelo BERT, capturando as nuances de significado entre as interpretações. [Zhang et al. \(2019\)](#) introduziram o BERTScore como uma métrica que aproveita embeddings contextuais, mostrando correlação superior com avaliações humanas em tarefas de geração de texto, como tradução automática e sumarização.
    1. **Geração de Embeddings:** Cada frase, tanto de referência quanto de candidato, é transformada em embeddings utilizando um modelo pré-treinado BERT, resultando em vetores de alta dimensionalidade que capturam o significado semântico das palavras.
    2. **Cálculo da Similaridade:** Para cada par de textos, a similaridade de cada token do texto candidato com os tokens da referência é calculada usando a Similaridade do Cosseno entre seus embeddings.
    3. **Pontuação Final:** O BERTScore é calculado considerando a média das maiores similaridades entre os tokens dos textos, balanceando precisão, revocação e F1 para representar a similaridade semântica geral entre as frases geradas e as de referência.

Matematicamente, o BERTScore pode ser interpretado como uma média ponderada das similaridades entre tokens, onde tokens semanticamente próximos no espaço de embeddings contribuem mais para o score final, proporcionando uma avaliação robusta da proximidade semântica entre textos.

# 5

## Resultados e Discussões

Neste capítulo, discutimos os resultados obtidos do modelo BERTopic e das métricas de comparação entre textos interpretados por humanos e textos analisados por um modelo de linguagem (LLM).

### 5.1 Dados Coletados e Preparação para Análise

O conjunto de dados foi delimitado ao número de postagens obtidas dentro de um intervalo de 5 minutos a partir do momento da requisição enviada para a API por perfil. Foram coletadas postagens de seis perfis feministas no Instagram, totalizando **11.582 postagens**. Após o processo de pré-processamento, que incluiu a limpeza de dados e a remoção de strings vazias, o conjunto final consistiu em **11.070 postagens** preparadas para análise. Cada *string* representava uma postagem, contendo as informações da legenda, textos presentes nas imagens e hashtags associadas, formando o corpus utilizado nas técnicas de modelagem de tópicos.

### 5.2 Resultados da Modelagem de Tópicos com BERTopic

A aplicação da técnica BERTopic resultou na identificação de **90 tópicos** distintos, cada um representado por 10 palavras-chave, calculadas utilizando o método `ClassTfidfTransformer` (TF-IDF). As palavras-chave identificadas são os termos mais relevantes dentro de cada grupo de postagens, refletindo os principais conceitos discutidos em cada tópico.

A Figura 5.1 apresenta os **Topic Word Scores** para os 12 tópicos mais representativos, ou seja, aqueles com maior volume de postagens associadas. O **Topic Word Score** é uma métrica que indica o grau de importância de cada palavra dentro do tópico em relação ao corpus total, com base no cálculo do TF-IDF. Quanto maior o score, mais relevante é a palavra para a definição daquele tópico em particular. Na imagem, estão listadas as palavras mais relevantes

para cada um dos 12 tópicos, ordenadas de acordo com sua importância relativa dentro de cada grupo.

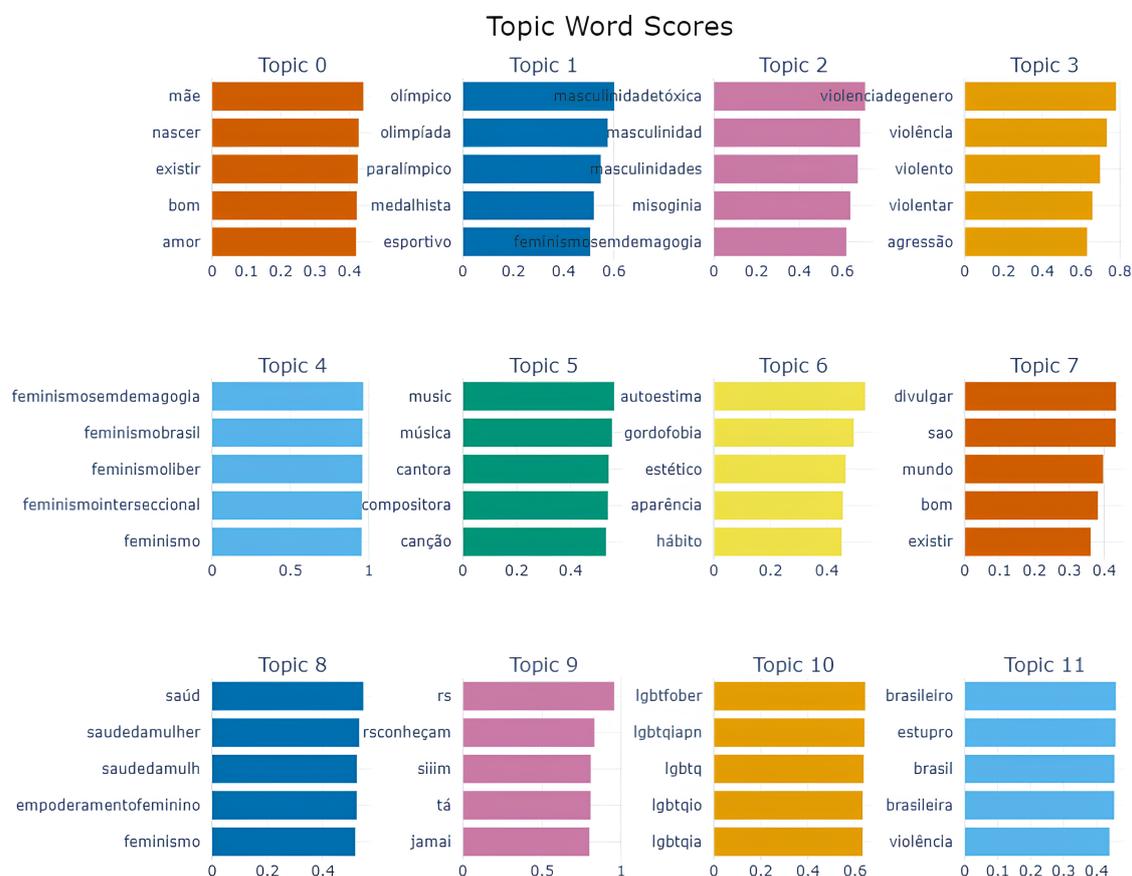


Figura 5.1: Topic Word Scores dos 12 principais tópicos.  
 Fonte: Thalia Barbosa Marques de Almeida. Acervo Pessoal.

Esses tópicos contêm o maior número de postagens, indicando que eles concentram os temas mais amplamente discutidos no conjunto de dados analisado.

### 5.2.1 Análise do Tópico -1: Postagens Atípicas ou Outliers

Durante a modelagem de tópicos com BERTopic, foi identificado um tópico especial denominado **Tópico -1**. Esse tópico agrupa postagens que não se encaixaram claramente em nenhum dos demais, sendo considerado um conjunto de *outliers* ou postagens atípicas. O **Tópico -1** contém um total de **5.373 documentos**, tornando-se o tópico com a maior quantidade de conteúdo.

Embora essas postagens não tenham sido agrupadas com os tópicos mais homogêneos, o **Tópico -1** possui relevância e não deve ser desconsiderado. A Figura 5.2 exibe uma nuvem de palavras que destaca as principais palavras associadas ao **Tópico -1**, onde os termos mais frequentes e representativos aparecem em maior destaque. Isso indica que, apesar de sua natureza

fragmentada, essas postagens podem abordar temas importantes para a compreensão ampla dos tópicos feministas.



Figura 5.2: Nuvem de palavras com as principais palavras associadas ao Tópico -1.  
Fonte: Thalia Barbosa Marques de Almeida. Acervo Pessoal.

### 5.3 Comparação das Interpretações Humanas e do LLM LLaMA

Como descrito no capítulo anterior, para validar a interpretação dos tópicos foi conduzido um experimento de interpretação com a participação de respondentes via Google Forms. O formulário ficou aberto ao público por 3 semanas, obtendo um total de **38 respostas**.

Para avaliar a consistência entre as interpretações realizadas pelos participantes humanos e as geradas automaticamente pelo modelo de linguagem LLaMA, foram utilizadas quatro métricas principais de comparação. As métricas selecionadas englobam tanto aspectos léxicos e sintáticos quanto aspectos semânticos do texto. As métricas **BLEU** e **ROUGE** analisam a similaridade a partir de uma perspectiva léxica, enquanto as métricas **Similaridade de Cosseno** e **BERTScore** consideram a proximidade semântica entre as interpretações. Cada uma dessas métricas proporciona insights únicos sobre os textos gerados, permitindo uma análise multidimensional das semelhanças e diferenças entre as interpretações humanas e a do modelo LLaMA.

ID	Saída LLM	Métricas			
		BLEU	ROUGE	Cosine Similarity	BERTScore
		Média   DP	Média   DP	Média   DP	Média   DP
0	Feminismo e Aborto	1.2750e-231   0.0	0.0853   0.1013	0.0364   0.0967	0.7256   0.0344
1	Maternidade	0.0000   0.0	0.3138   0.2745	0.3769   0.2820	0.7869   0.0855
2	Esporte Olímpico	0.0000   0.0	0.1352   0.1550	0.0861   0.1330	0.6803   0.0416
3	Masculinidades e feminismo	1.0929e-231   0.0	0.0970   0.1496	0.0000   0.0000	0.7361   0.0496
4	Violência Doméstica	0.0000   0.0	0.3858   0.3272	0.3076   0.3263	0.7657   0.0923
5	Maquiagem e Aparência	1.2750e-231   0.0	0.0531   0.1254	0.0215   0.0757	0.6900   0.0320
6	Feminismo	0.0000   0.0	0.0417   0.1730	0.0481   0.1844	0.7506   0.0625
7	Luta contra a homofobia e discriminação LGBT	8.8504e-232   0.0	0.1257   0.1455	0.0745   0.1374	0.6835   0.0685
8	Discriminação racial e racismo	8.8142e-232   0.0	0.2036   0.1476	0.1979   0.1535	0.7830   0.0567
9	Transtornos mentais e Psicologia	1.2543e-232   0.0	0.0525   0.0961	0.0179   0.0534	0.7279   0.0403

Tabela 5.1: A tabela compara as interpretações do modelo LLM e de participantes humanos, mostrando médias e desvios-padrão das métricas de avaliação.

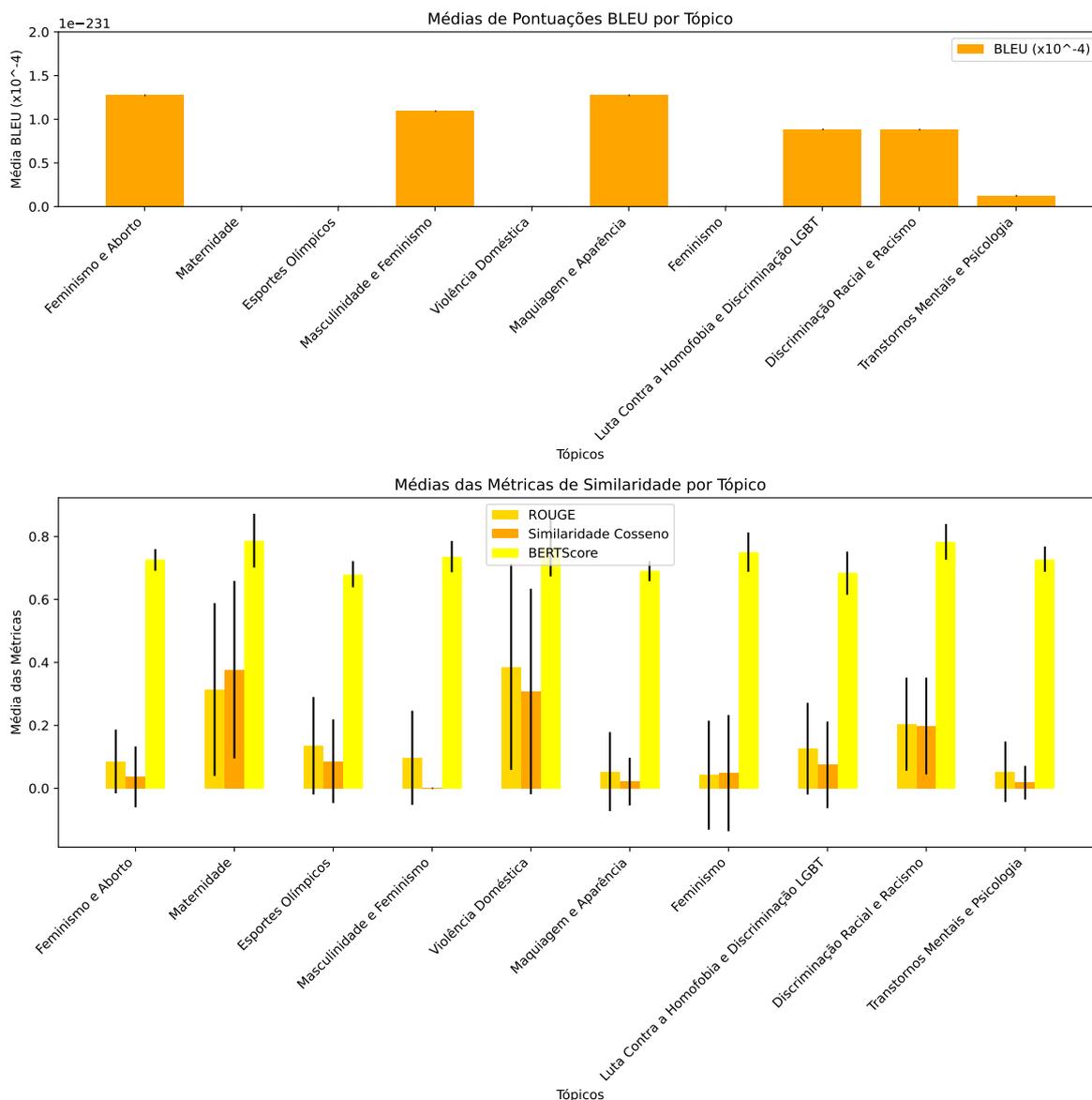


Figura 5.3: Médias e variabilidade das métricas de coerência e similaridade por tópico. As barras representam as médias das métricas BLEU, ROUGE, Cosine Similarity e BERTScore para cada tema, com barras de erro indicando o desvio padrão, facilitando a comparação entre os diferentes tópicos.

**Fonte:** Thalia Barbosa Marques de Almeida. Acervo Pessoal.

A análise das métricas de coerência e similaridade para cada tópico permitiu uma comparação abrangente do desempenho do modelo em temas variados. Na Figura 5.3, são apresentadas as médias das métricas BLEU, ROUGE, Cosine Similarity e BERTScore, com barras de erro indicando o desvio padrão para cada tópico avaliado. Essa visualização facilita a compreensão das diferenças e variabilidade dos resultados entre os temas, destacando onde o modelo apresenta maior ou menor consistência.

### 5.3.1 Similaridade entre as respostas humanas com o LLaMA

Em todas as métricas aplicadas, um valor próximo de 1.0 indica uma alta similaridade entre as respostas, sugerindo que o modelo conseguiu reproduzir o mesmo significado ou estrutura da resposta humana. Por outro lado, um valor próximo de 0.0 representa uma baixa similaridade, indicando que a resposta gerada pelo modelo diverge significativamente da resposta humana em conteúdo ou contexto.

#### Similaridade lexical

Ao observar os valores da Tabela 5.1, podemos analisar as métricas que refletem a **similaridade lexical** entre os textos gerados pelo LLM e as respostas humanas. Os valores médios de **BLEU** são extremamente baixos para todos os tópicos, próximos de zero, indicando que há pouca coincidência literal entre as saídas do modelo e as respostas humanas. Isso sugere que o LLM gera respostas que, embora possam capturar o significado geral, não utilizam as mesmas palavras ou n-gramas presentes nas interpretações humanas. Em outras palavras, o BLEU baixo indica que o modelo não está reproduzindo o vocabulário ou a estrutura exata das respostas humanas.

Por outro lado, os valores médios de **ROUGE** mostram uma variação mais ampla entre os tópicos. Em alguns casos, como *Violência Doméstica* (0.3858) e *Maternidade* (0.3138), os valores de ROUGE são significativamente mais altos, sugerindo uma maior sobreposição lexical com as respostas humanas. Isso implica que, nesses tópicos específicos, o LLM conseguiu reproduzir termos ou expressões-chave que aparecem nas interpretações humanas, resultando em uma maior similaridade de palavras. A diferença de comportamento entre BLEU e ROUGE pode ser atribuída ao fato de que ROUGE se concentra mais em **recall**, ou seja, mede quantas palavras ou frases das respostas humanas estão presentes nas respostas do modelo, independentemente de ordem exata ou precisão dos n-gramas.

Em contraste, outros tópicos, como *Maquiagem e Aparência* (0.0531) e *Transtornos Mentais e Psicologia* (0.0525), apresentam valores de ROUGE muito baixos, o que indica pouca sobreposição lexical com as respostas humanas. Esses valores baixos de ROUGE, combinados com o BLEU também próximo de zero, sugerem que, nesses tópicos, o LLM utiliza um vocabulário consideravelmente diferente do empregado pelos participantes humanos, resultando em uma menor similaridade textual.

#### Similaridade semântica

Os valores médios de **Similaridade de Cosseno** são relativamente baixos na maioria dos tópicos, variando de 0.02 a 0.37. Destacam-se os tópicos *Maternidade* (0.3769) e *Violência Doméstica* (0.3076), onde os valores são ligeiramente mais altos. A Similaridade de Cosseno mede a similaridade com base em palavras-chave, utilizando representações vetoriais TF-IDF. Valores baixos de Similaridade de Cosseno sugerem que o LLM usa palavras diferentes das respostas

humanas, o que resulta em uma baixa similaridade quando comparado no nível de palavras isoladas. Mesmo nos tópicos com valores um pouco mais elevados, isso pode indicar apenas uma coincidência superficial de termos, sem necessariamente capturar o significado completo.

Por outro lado, os valores de **BERTScore** são consistentemente mais altos, variando de 0.68 a 0.78, com destaque novamente nos tópicos *Maternidade* (0.7869) e *Violência Doméstica* (0.7657). O BERTScore utiliza embeddings contextuais que capturam a semântica dos textos, permitindo que ele reconheça similaridades de significado mesmo quando não há coincidência exata de palavras. O BERTScore com valores altos indica que o LLM está se aproximando do conteúdo semântico das respostas humanas, ainda que use diferentes expressões para transmitir a mesma ideia. A figura 5.4 ilustra os tópicos em que as métricas semânticas capturaram maiores similaridades entre os textos de referência e os textos do LLM.

A discrepância entre Similaridade de Cosseno e BERTScore, pode ser atribuída às diferenças na forma como cada métrica mede a similaridade semântica. A Similaridade de Cosseno depende da presença de termos específicos e, portanto, é menos eficaz em capturar o significado semântico completo quando o vocabulário do LLM difere das respostas humanas. Já o BERTScore, ao utilizar embeddings contextuais, consegue medir a similaridade de significado de maneira muito mais robusta, capturando nuances semânticas mesmo com variações de vocabulário.

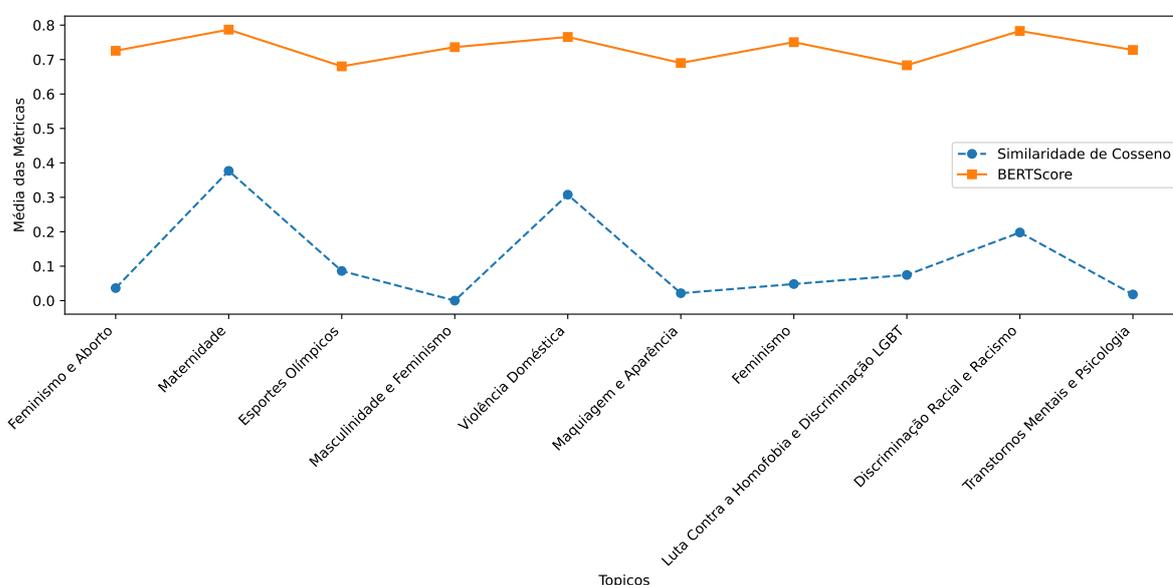


Figura 5.4: Comparação de similaridade semântica por tópico (Similaridade de Cosseno vs BERTScore).

**Fonte:** Thalia Barbosa Marques de Almeida. Acervo Pessoal.

Em resumo, a partir da visão geral dos resultados ilustrados na Figura 5.5, podemos concluir que os textos referenciais humanos utilizam palavras distintas (evidenciado pelo valor médio próximo de zero de BLEU) ou palavras-chave iguais ou similares (evidenciado pelos valores de ROUGE e Similaridade de Cosseno), mas com significados muito próximos (evidenciado pela consistência e altos valores de BERTScore) aos textos gerados pelo LLaMA.

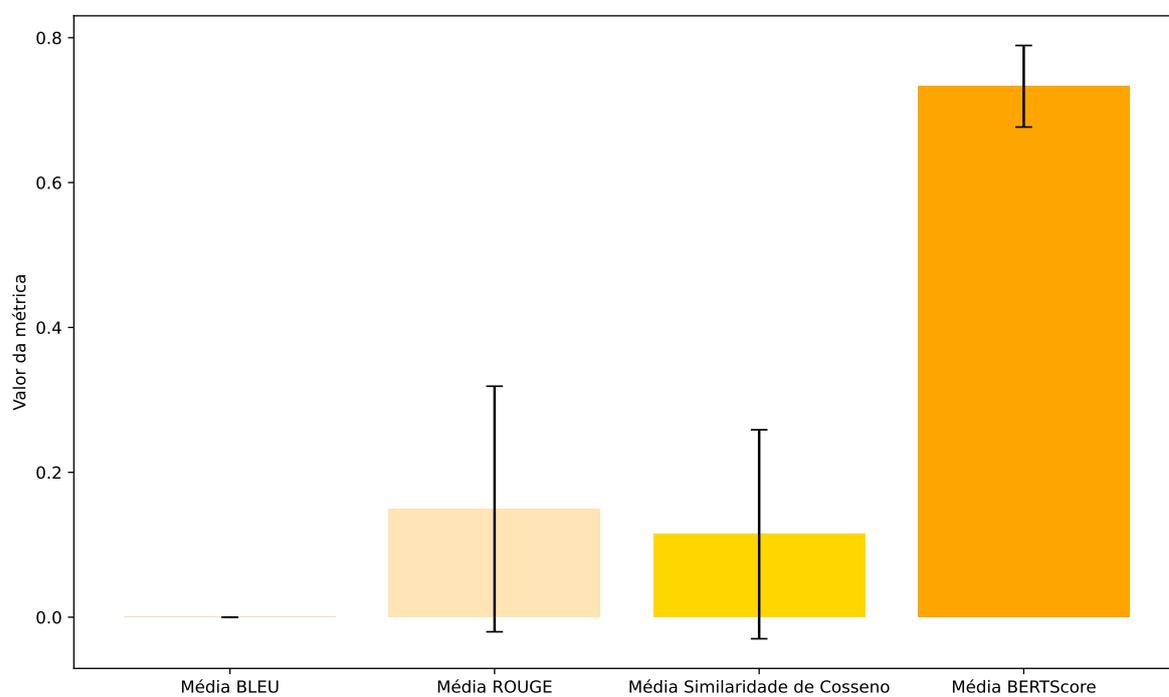


Figura 5.5: Médias gerais das métricas com barras de erro que representam o desvio padrão.

**Fonte:** Thalia Barbosa Marques de Almeida. Acervo Pessoal.

# 6

## Conclusão

O presente estudo teve como objetivo identificar e interpretar os temas abordados por perfis feministas brasileiros no Instagram, aplicando técnicas de Processamento de Linguagem Natural (PLN) e modelagem de tópicos para analisar as discussões predominantes nesses espaços. A combinação entre a técnica BERTopic e o modelo de linguagem de grande escala LLaMA permitiu uma análise robusta e interpretativa dos temas, contribuindo para a compreensão das narrativas feministas contemporâneas na plataforma.

A modelagem de tópicos com BERTopic resultou na identificação de 90 temas distintos, destacando-se doze tópicos principais, como violência doméstica, direitos reprodutivos, saúde mental, maternidade, autonomia do corpo e igualdade de gênero. Esses temas refletem debates centrais do feminismo no Brasil, país que, nos últimos anos, tem enfrentado discussões acirradas sobre direitos das mulheres, especialmente em relação à violência de gênero e aos direitos sobre o próprio corpo. A ênfase no tema da violência doméstica, por exemplo, ressalta a importância dessa questão para as feministas brasileiras, dado o cenário nacional de altas taxas de violência contra a mulher, o que torna a conscientização sobre o problema e a exigência de políticas públicas essenciais. Outros temas, como os direitos reprodutivos e a saúde mental, também se destacam como tópicos de alta relevância, considerando as recentes discussões sobre acesso a serviços de saúde e o impacto da saúde mental na vida das mulheres.

A análise comparativa entre as interpretações humanas e as respostas geradas pelo modelo LLaMA mostrou que, embora a similaridade lexical (avaliada por métricas como BLEU e ROUGE) tenha sido baixa, o BERTScore indicou uma alta similaridade semântica. Essa diferença é esperada e até mesmo desejável no contexto do nosso experimento, uma vez que o foco estava na interpretação dos significados, não na replicação exata de termos usados pelos participantes humanos. A alta performance no BERTScore sugere que o LLaMA foi capaz de capturar a essência dos temas discutidos, interpretando o conteúdo com coerência em relação aos significados principais. Portanto, com base nos resultados do BERTScore, conclui-se que

o modelo LLaMA demonstra uma capacidade relevante para interpretar e nomear tópicos representados por palavras-chave, sugerindo um potencial para substituir a intervenção humana nessa tarefa.

Diante disso, a hipótese alternativa (H1), que prevê a capacidade das técnicas de modelagem de tópicos combinadas com LLMs em identificar e interpretar temas feministas no Instagram, foi corroborada, especialmente no aspecto de similaridade semântica. Os resultados indicam que modelos de linguagem de grande escala, como o LLaMA, são apropriados para análises onde o contexto e o significado dos temas são mais importantes do que a correspondência lexical exata, posicionando-se como ferramentas promissoras para a interpretação de dados textuais em plataformas de mídias sociais.

A combinação do BERTopic com um modelo LLM para a interpretação de tópicos mostrou-se uma abordagem promissora para a análise de grandes volumes de dados textuais em redes sociais. Este estudo contribui para os estudos sociais em plataformas digitais, evidenciando a importância das discussões feministas no Instagram e destacando a necessidade de avanços contínuos em PLN para melhorar ainda mais a capacidade de análise semântica desses conteúdos.

## 6.1 Trabalhos futuros

Com o avanço dos modelos de linguagem de grande escala (LLMs), torna-se possível simplificar a análise de temas em mídias sociais, potencialmente dispensando etapas tradicionais de modelagem de tópicos. Estudos futuros poderiam explorar a viabilidade de utilizar exclusivamente LLMs para identificar e interpretar temas, sem o apoio de técnicas adicionais como o BERTopic. Essa abordagem poderia oferecer uma solução mais ágil e integrada, aproveitando o potencial dos LLMs para gerar interpretações semânticas profundas diretamente dos dados coletados, facilitando a análise de discussões sociais online.

Contudo, a aplicação dessa metodologia apresenta desafios, especialmente devido à complexidade e imprevisibilidade das respostas dos LLMs. Futuros estudos poderiam investigar estratégias para aumentar a confiabilidade e consistência dos temas extraídos, reduzindo variações indesejadas. Técnicas como engenharia de prompt e ajuste fino podem ser fundamentais para estruturar os prompts ou adaptar o modelo, guiando a interpretação dos temas com maior precisão e alinhamento aos objetivos da análise.

## Referências bibliográficas

- Amin, M. H., Mohamed, E. K., and Elragal, A. (2020). Corporate disclosure via social media: a data science approach. *Online Information Review*, 44(1):278–298.
- Bahja, M. and Safdar, G. A. (2020). Unlink the link between covid-19 and 5g networks: an nlp and sna based approach. *Ieee Access*, 8:209127–209137.
- Banet-Weiser, S. (2018). *Empowered: Popular Feminism and Popular Misogyny*. Duke University Press.
- Bauer, B., Norel, R., Leow, A., Rached, Z. A., Wen, B., and Cecchi, G. (2024). Using large language models to understand suicidality in a social media–based taxonomy of mental health disorders: Linguistic analysis of reddit posts. *JMIR mental health*, 11:e57234.
- Bérubé, M., Tang, T.-U., Fortin, F., Ozalp, S., Williams, M. L., and Burnap, P. (2020). Social media forensics applied to assessment of post–critical incident social reaction: The case of the 2017 manchester arena terrorist attack. *Forensic science international*, 313:110364.
- Blei, D. M. and Lafferty, J. D. (2007). A correlated topic model of science.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.
- Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. *Proceedings of GSCL*, 30:31–40.
- Boyd-Graber, J., Hu, Y., Mimno, D., et al. (2017). Applications of topic models. *Foundations and Trends® in Information Retrieval*, 11(2-3):143–296.
- Brown, N. M. (2019). Methodological cyborg as black feminist technology: constructing the social self using computational digital autoethnography and social media. *Cultural Studies Critical Methodologies*, 19(1):55–67.
- Brown, P. F., Della Pietra, V. J., Desouza, P. V., Lai, J. C., and Mercer, R. L. (1992). Class-based n-gram models of natural language. *Computational linguistics*, 18(4):467–480.

- Cai, M., Shah, N., Li, J., Chen, W.-H., Cuomo, R. E., Obradovich, N., and Mackey, T. K. (2020). Identification and characterization of tweets related to the 2015 indiana hiv outbreak: a retrospective infoveillance study. *Plos one*, 15(8):e0235150.
- Chowdhury, G. G. (2010). *Introduction to modern information retrieval*. Facet publishing.
- Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Diógenes, E. M. N., Rocha, M. C. J., and Brabo, T. S. A. M. (2015). Os movimentos feministas brasileiros na luta pelos direitos das mulheres. In Brabo, T. S. A. M., editor, *Educação, mulheres, gênero e violência*, pages 305–322. Oficina Universitária; Cultura Acadêmica, Marília; São Paulo.
- DuBois, E. (1975). *The radicalism of the woman suffrage movement: Notes toward the reconstruction of nineteenth-century feminism*. Ann Arbor, MI: Michigan Publishing, University of Michigan Library.
- Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.
- Hacker, J., Vom Brocke, J., Handali, J., Otto, M., and Schneider, J. (2020). Virtually in this together—how web-conferencing systems enabled a new virtual togetherness during the covid-19 crisis. *European Journal of Information Systems*, 29(5):563–584.
- Harris, Z. (1954). Distributional structure.
- Hemmatian, B., Sloman, S. J., Cohen Priva, U., and Sloman, S. A. (2019). Think of the consequences: A decade of discourse about same-sex marriage. *Behavior research methods*, 51:1565–1585.
- Hoyle, A., Goel, P., Hian-Cheong, A., Peskov, D., Boyd-Graber, J., and Resnik, P. (2021). Is automated topic model evaluation broken? the incoherence of coherence. *Advances in neural information processing systems*, 34:2018–2033.
- Ibrahim, N. F. and Wang, X. (2019). A text analytics approach for online retailing service improvement: Evidence from twitter. *Decision Support Systems*, 121:37–50.
- Joo, S., Lu, K., and Lee, T. (2020). Analysis of content topics, user engagement and library factors in public library social media based on text mining. *Online information review*, 44(1):258–277.
- Kenton, J. D. M.-W. C. and Toutanova, L. K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of naacL-HLT*, volume 1, page 2. Minneapolis, Minnesota.

- Kirilenko, A. and Stepchenkova, S. (2024). Automated topic analysis with large language models. In *ENTER e-Tourism Conference*, pages 29–34. Springer.
- Kurten, S. and Beullens, K. (2021). # coronavirus: Monitoring the belgian twitter discourse on the severe acute respiratory syndrome coronavirus 2 pandemic. *Cyberpsychology, Behavior, and Social Networking*, 24(2):117–122.
- Kwon, O. H., Vu, K., Bhargava, N., Radaideh, M. I., Cooper, J., Joynt, V., and Radaideh, M. I. (2024). Sentiment analysis of the united states public support of nuclear power on social media using large language models. *Renewable and Sustainable Energy Reviews*, 200:114570.
- Laureate, C. D. P., Buntine, W., and Linger, H. (2023). A systematic review of the use of topic models for short text social media analysis. *Artificial Intelligence Review*, 56(12):14223–14255.
- LESFEM - Monitor de Femicídios do Brasil (2024). Monitor de feminicídios do brasil revela aumento alarmante e estatísticas preocupantes. Acessado em: 3 de novembro de 2024.
- Liang, B., Wang, Y., and Tsou, M.-H. (2019). A “fitness” theme may mitigate regional prevalence of overweight and obesity: Evidence from google search and tweets. *Journal of Health Communication*, 24(9):683–692.
- Lin, C.-Y. (2004). ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.
- Manning, C. and Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT press.
- Mazarura, J. and de Waal, A. (2016). A comparison of the performance of latent dirichlet allocation and the dirichlet multinomial mixture model on short text. In *2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)*, pages 1–6.
- Mendes, K., Ringrose, J., and Keller, J. (2018). #metoo and the promise and pitfalls of challenging rape culture through digital feminist activism. *European Journal of Women’s Studies*, 25(2):236–246.
- Mikolov, T. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 3781.
- Muthusami, R., Mani Kandan, N., Saritha, K., Narenthiran, B., Nagaprasad, N., and Ramaswamy, K. (2024). Investigating topic modeling techniques through evaluation of topics discovered in short texts data across diverse domains. *Scientific Reports*, 14(1):12003.

- Nobles, A. L., Leas, E. C., Latkin, C. A., Dredze, M., Strathdee, S. A., and Ayers, J. W. (2020). # hiv: alignment of hiv-related visual content on instagram with public health priorities in the us. *AIDS and Behavior*, 24:2045–2053.
- Okon, E., Rachakonda, V., Hong, H. J., Callison-Burch, C., and Lipoff, J. B. (2020). Natural language processing of reddit data to evaluate dermatology patient experiences and therapeutics. *Journal of the American Academy of Dermatology*, 83(3):803–808.
- Paasonen, S. (2011). Revisiting cyberfeminism.
- Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. In Isabelle, P., Charniak, E., and Lin, D., editors, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.
- Parsing, C. (2009). Speech and language processing. *Power Point Slides*.
- Radford, A. (2018). Improving language understanding by generative pre-training.
- Rahutomo, F., Kitasuka, T., Aritsugi, M., et al. (2012). Semantic cosine similarity. In *The 7th international student conference on advanced science and technology ICAST*, volume 4, page 1. University of Seoul South Korea.
- Rijcken, E., Scheepers, F., Zervanou, K., Spruit, M., Mosteiro, P., and Kaymak, U. (2023). Towards interpreting topic models with chatgpt. In *The 20th World Congress of the International Fuzzy Systems Association*.
- Saavedra, J. M. (2019). Una nueva ola feminista, más allá de# metoo: Irrupción, legado y desafíos. *Políticas Públicas para la Equidad*, 2.
- Saeki, T., Maiti, S., Takamichi, S., Watanabe, S., and Saruwatari, H. (2024). Speechbertscore: Reference-aware automatic evaluation of speech generation leveraging nlp evaluation metrics. *arXiv preprint arXiv:2401.16812*.
- Sia, S., Dalmia, A., and Mielke, S. J. (2020). Tired of topic models? clusters of pretrained word embeddings make for fast and good topics too! *arXiv preprint arXiv:2004.14914*.
- Statista (2024). Instagram: number of global users 2020-2025. Acessado em: 3 de novembro de 2024.
- Supremo Tribunal Federal (2023). Supremo realiza importante votação sobre direitos fundamentais. Acessado em: 3 de novembro de 2024.

- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al. (2023). Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.
- Vachhani, S. J. (2024). Networked feminism in a digital age—mobilizing vulnerability and reconfiguring feminist politics in digital activism. *Gender, Work & Organization*, 31(3):1031–1048.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *advances in neural information processing systems*, 30(2017).
- White, D. M. (1950). The “gate keeper”: A case study in the selection of news. *Journalism Quarterly*, 27(4):383–390.
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., and Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382*.
- Xu, S. and Xiong, Y. (2020). Setting socially mediated engagement parameters: A topic modeling and text analytic approach to examining polarized discourses on gillette’s campaign. *Public Relations Review*, 46(5):101959.
- Young, T., Hazarika, D., Poria, S., and Cambria, E. (2018). Recent trends in deep learning based natural language processing. *iee Computational intelligence magazine*, 13(3):55–75.
- Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., and Artzi, Y. (2019). Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.