

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
**FACULDADE DE LETRAS**  
**CURSO DE ESPECIALIZAÇÃO EM LINGUAGEM JURÍDICA**

**Maria Clara de Neto Sales Oliveira**

**IMPOLIDEZ NO AMBIENTE DIGITAL: Medidas Implementadas pelo *Instagram* para  
Evitar o Discurso de Ódio na Rede**

**BELO HORIZONTE - MG**

**2024**

**Maria Clara de Neto Sales Oliveira**

**IMPOLIDEZ NO AMBIENTE DIGITAL: Medidas Implementadas pelo *Instagram* para Evitar o Discurso de Ódio na Rede**

Trabalho de Conclusão de Curso, no formato de artigo científico, apresentado ao Colegiado do Curso de Especialização em Linguagem Jurídica da Universidade Federal de Minas Gerais, como requisito para a obtenção do Título de Especialista em Linguagem Jurídica.

**Orientador: Prof.<sup>a</sup> Doutora Monique Vieira Miranda.**

**BELO HORIZONTE – MG**

**2024**



UNIVERSIDADE FEDERAL DE MINAS GERAIS

## **ATA**

FALE - SECRETARIA GERAL

### ATA DA DEFESA DE TRABALHO DE CONCLUSÃO DE CURSO

Nome do(a) aluno(a): Maria Clara de Neto Sales Oliveira

Matrícula: 2023654895

Às 10:45 horas do dia 15 de junho de 2024, reuniu-se, na Faculdade de Letras da UFMG, a Comissão Examinadora indicada pela Coordenação do Curso de Especialização em Linguagem Jurídica, para avaliar, em exame final, o trabalho intitulado "IMPOLIDEZ NO AMBIENTE DIGITAL: Medidas Implementadas pelo Instagram para Evitar o Discurso de Ódio na Rede", como requisito final para obtenção do Grau de Especialista em Linguagem Jurídica. Abrindo a sessão, a Comissão Examinadora, após dar conhecimento aos presentes do teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa da candidata. Em seguida, a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações:

Prof. Dr. Luiz Francisco Dias indicou a aprovação da candidata;

Profa. Dra. Fabiana Meireles de Oliveira indicou a aprovação da candidata.

Pelas indicações, a candidata foi aprovada.

Nota: 100

O resultado final foi comunicado publicamente à candidata pela banca. Nada mais havendo a tratar, a banca encerrou a sessão, da qual foi lavrada a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

O trabalho atende aos requisitos do Curso de Especialização em Linguagem Jurídica.



Documento assinado eletronicamente por **Fabiana Meireles de Oliveira, Usuário Externo**, em 19/06/2024, às 16:22, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Luiz Francisco Dias, Servidor(a)**, em 20/06/2024, às 12:10, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

A autenticidade deste documento pode ser conferida no site



[https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **3314155** e o código CRC **F517D96D**.

## RESUMO

Este artigo examina a eficácia das estratégias de moderação de conteúdo nas plataformas de mídia social, especificamente a do Instagram, destacando suas limitações. Para alcançar este objetivo, elaborou-se um estudo bibliográfico do tipo integrativo, o qual envolveu um levantamento dos textos acadêmicos (Araújo, 2022; Caldeira, 2022; Culpeper, 2021; Cunha, 2019; Cunha e Oliveira, 2020; Oliveira e Miranda, 2020), pesquisas na internet acerca das políticas da rede social estudada, bem como o livro de direito constitucional do autor Bernardo Gonçalves (Gonçalves, 2010). Com base na teoria de impolidez linguística (Culpeper, 2021) e no conceito de discurso de ódio (Culpeper, 2021) e do próprio conceito de discurso de ódio elaborado pelo Instagram, foi feita uma análise dos mecanismos utilizados pelo Instagram para mitigar os discursos ofensivos e de ódio, trazendo, a título de exemplificação, uma sucinta análise de comentários em postagens do jogador de futebol Vini Júnior, a fim de se verificar se esses mecanismos são eficazes e/ou cerceiam a liberdade de expressão dos usuários. Com base nesse estudo, pôde ser constatado que apesar de existirem medidas, como algoritmos de detecção e equipes de revisão, objetivando reduzir assuntos ofensivos, verificou-se a existência de dificuldades para filtrar completamente o conteúdo ofensivo, como exemplificado no caso envolvendo o jogador mencionado. A análise teórica revelou ainda desafios significativos, incluindo a sobrecarga de conteúdo e a dificuldade em distinguir entre discurso legítimo e ofensivo. Essas deficiências apontam para a necessidade de aprimoramentos nas estratégias de moderação. Concluiu-se que, embora a plataforma de mídia social do Instagram tenha tentado abordar e moderar a disseminação de conteúdo prejudicial, ainda existem lacunas na capacidade de garantir um ambiente online seguro e livre de discursos ofensivos.

**Palavras-chave:** Discurso de ódio, Impolidez linguística, Moderação de conteúdo, Instagram, Conteúdo ofensivo, Liberdade de expressão;

## ABSTRACT

This article examines the effectiveness of content moderation strategies on social media platforms, specifically Instagram, highlighting their limitations. To achieve this objective, an integrative bibliographic study was prepared, which involved a survey of academic texts (Araújo, 2022; Caldeira, 2022; Culpeper, 2021; Cunha, 2019; Cunha and Oliveira, 2020; Oliveira and Miranda, 2020), internet research on the policies of the social network studied, as well as the book on constitutional law by author Bernardo Gonçalves (Gonçalves, 2010). Based on the theory of linguistic impoliteness (Culpeper, 2021) and the concept of hate speech (Culpeper, 2021) and the definition of hate speech developed by Instagram, an analysis was carried out on the mechanisms used by Instagram to mitigate offensive and hate speech, bringing, by way of example, a succinct analysis of comments on posts by football player Vini Júnior, in order to verify whether these mechanisms are effective and/or restrict users' freedom of expression. Based on this study, it was found that although there are measures, such as detection algorithms and review teams, aimed at reducing offensive subjects, there were

difficulties in completely filtering offensive content, as exemplified in the case involving the aforementioned player. The theoretical analysis also revealed significant challenges, including content overload and the difficulty in distinguishing between legitimate and offensive speech. These deficiencies point to the need for improvements in moderation strategies. It concluded that although the Instagram social media platform has attempted to address and moderate the spread of harmful content, there are still gaps in its ability to ensure a safe online environment free from offensive speech.

**Keywords:** Hate speech, Linguistic impoliteness, Content moderation, Instagram, Offensive content, Freedom of expression;

## **SUMÁRIO**

<b>1. INTRODUÇÃO .....</b>	<b>7</b>
<b>1.1 Objetivo, hipótese e conteúdo do texto .....</b>	<b>7</b>
<b>2. IMPOLIDEZ LINGUÍSTICA E DISCURSO DE ÓDIO NA ERA DIGITAL .....</b>	<b>10</b>
<b>3. MEDIDAS DE MODERAÇÃO DO INSTAGRAM CONTRA O DISCURSO DE ÓDIO ..</b>	<b>15</b>
<b>4. ANÁLISE DAS MEDIDAS DO INSTAGRAM: MODERAÇÃO E LIBERDADE DE EXPRESSÃO.....</b>	<b>18</b>
<b>5. CONSIDERAÇÕES FINAIS .....</b>	<b>22</b>
<b>6. ANEXOS .....</b>	<b>27</b>
<b>Anexo I – Print de Publicação no Instagram do Jogador Vinicius Junior.....</b>	<b>27</b>
<b>Anexo II – Print de Publicação no Instagram do Jogador Vinicius Junior .....</b>	<b>28</b>
<b>Anexo III – Print de Publicação no Instagram do Jogador Vinicius Junior .....</b>	<b>29</b>
<b>Anexo IV – Print de Publicação no Instagram do Jogador Vinicius Junior .....</b>	<b>30</b>
<b>Anexo V – Print de Publicação no Instagram do Jogador Vinicius Junior .....</b>	<b>31</b>
<b>Anexo VI – Print de Publicação no Instagram do Jogador Vinicius Junior .....</b>	<b>32</b>
<b>Anexo VII – Print de Publicação no Instagram do Jogador Vinicius Junior .....</b>	<b>33</b>
<b>Anexo VIII – Print de Publicação no Instagram do Jogador Vinicius Junior.....</b>	<b>34</b>
<b>Anexo IX – Print de Publicação no Instagram do Jogador Vinicius Junior.....</b>	<b>35</b>

## 1. INTRODUÇÃO

### 1.1 Objetivo, hipótese e conteúdo do texto

A história da internet é marcada por uma evolução constante, desde suas origens nos anos 1960, quando foi concebida como uma rede de comunicação militar, até se tornar a vasta rede global que conhecemos hoje. Com o advento da *World Wide Web*<sup>1</sup> na década de 1990, a internet passou a se popularizar entre o público em geral, permitindo o compartilhamento de informações e o surgimento das primeiras redes sociais.

Foi nesse contexto que o Instagram emergiu como uma plataforma de compartilhamento de fotos e vídeos, lançada em 2010 e adquirida pelo Facebook em 2012. Ao longo dos anos, essa plataforma de mídia social ganhou uma enorme base de usuários e se tornou uma das redes sociais mais populares do mundo, especialmente entre os jovens.

Assim como em outras épocas da história, em que as pessoas se reuniam em praças públicas ou outros espaços físicos para expressar suas opiniões e ideologias, o Instagram se tornou um novo palco virtual para diversas formas de expressão, incluindo o discurso de ódio. Esta plataforma, apesar de ser um espaço de interação e conexão positiva, também se tornou um ambiente propício para a disseminação de discursos prejudiciais e ofensivos. Isso porque o ambiente virtual cria a ideia de anonimato diante do distanciamento gerado pela tela e de impunidade entre os seus usuários, tornando-se necessário mecanismos como políticas de controle de postagens de cunho ofensivo.

Tal realidade provocou um desafio delicado a ser enfrentado pelas redes sociais, como o Instagram: moderar o discurso de ódio sem cercear a liberdade de expressão. Ao mesmo tempo em que é fundamental proteger os usuários contra abusos e violência online, também é necessário garantir que as plataformas não impeçam a livre troca de ideias e opiniões legítimas.

Assim, a questão do equilíbrio entre a moderação do conteúdo prejudicial e a preservação da liberdade de expressão torna-se central na análise das políticas e práticas adotadas por essa e outras redes sociais. A eficácia dessas medidas não apenas influencia a qualidade do ambiente digital, mas também reflete os valores e compromissos das empresas em relação aos direitos dos usuários e à segurança online.

Essa nova forma de interagir com o mundo tornou-se um meio em que ofensas e preconceitos passaram a acontecer de maneira recorrente e em grande número. Diante dessa

---

<sup>1</sup> Ambiente virtual que introduziu a multimídia em computadores conectados a servidores, permitindo a navegação por sítios interligados por links. Os navegadores possibilitam o acesso a documentos contendo texto, imagem, vídeo e gráficos em multimídia, de acordo com Fernanda Nahuz (1999).

frequência, as mídias sociais, como Facebook, X (antigo Twitter) e Instagram, criaram mecanismos para evitar que o discurso ódio esteja presente de maneira acentuada em suas plataformas. Nesse sentido, é importante que, de fato, esses mecanismos estejam cumprindo com o intento para o qual foram criados, a fim de que os direitos dos usuários não sejam tão ameaçados.

Nesse contexto, o presente artigo tem o objetivo de contribuir com o campo de estudos sobre a impolidez e o discurso de ódio no ambiente digital e sua relação com a liberdade de expressão ao responder à seguinte pergunta: Há moderação eficaz do discurso de ódio em comentários nas postagens do Instagram?

Cabe ressaltar que o enfoque dado a essa plataforma neste trabalho se justifica pelo fato de ela ter se tornado uma das formas de entretenimento favoritas dos brasileiros, sendo a segunda rede social mais utilizada, perdendo apenas para o WhatsApp.<sup>2</sup>

De forma mais específica, para atingir o objetivo geral desse artigo, foi realizado um levantamento bibliográfico sobre impolidez linguística e sobre discurso de ódio. Em seguida, realizou-se um estudo dos mecanismos que o Instagram utiliza para moderar discursos ofensivos e o discurso de ódio. Na sequência, foi feita uma sucinta análise de algumas postagens do jogador de futebol Vinicius José Paixão de Oliveira Júnior, mais conhecido como Vini Júnior, em seu Instagram pessoal, de forma a verificar se os mecanismos utilizados pela plataforma, a fim de moderar o conteúdo, são eficazes e se cerceiam ou não a liberdade de expressão do usuário, respondendo, desta feita, à pergunta levantada pela pesquisa.

A escolha de analisar especificamente o jogador se deu pela frequente incidência de notícias sobre ofensas direcionadas a ele em suas redes sociais. Tais ocorrências costumam gerar debates acalorados sobre essas ofensas constituírem ou não discurso de ódio. O fato de ele ser jogador de futebol o coloca como uma figura pública bastante conhecida, o que faz com que receba uma grande variedade de comentários, desde elogios até ataques pessoais, especialmente em suas redes sociais. Dado o grande interesse que o futebol desperta entre as pessoas, os debates em torno de jogadores famosos podem facilmente ultrapassar os limites e se transformar em discurso de ódio. Por isso, analisar os comentários dirigidos a esse jogador oferece uma oportunidade para entender como o Instagram lida com discursos ofensivos e se seus mecanismos de moderação são eficazes.

---

<sup>2</sup> Disponível em: <https://www.conversion.com.br/blog/redes-sociais/>. Acesso em 23 de abril de 2024.

A relevância dos ataques racistas direcionados ao jogador foi ainda mais enfatizada pela sua inclusão em uma nota oficial de imprensa emitida pelo Ministério das Relações Exteriores, de nº 119 e datada de março de 2024<sup>3</sup>. Tal menção oficial ressalta a gravidade e o impacto desses incidentes, bem como eleva o debate para além do âmbito esportivo: para uma questão de dimensão nacional e internacional. A inclusão desse tema em um comunicado oficial do governo sublinha a importância de enfrentar e combater o racismo no contexto do esporte e da sociedade em geral, inclusive, no meio virtual.

Nesse sentido, a hipótese formulada neste estudo sugere que os mecanismos de moderação implementados pelo Instagram não conseguem efetivamente mitigar a presença de discurso de ódio em suas postagens, resultando em uma moderação menos eficaz do que o desejado e proposto pela rede social. Concomitantemente, acredita-se que esses mecanismos não cerceariam a liberdade de expressão dos usuários.

Do ponto de vista metodológico, o conteúdo do texto foi construído, em sua parte teórica, com base em um estudo do tipo bibliográfico integrativo o qual envolveu um levantamento dos textos acadêmicos (Araújo, 2022; Caldeira, 2022; Culpeper, 2021; Cunha, 2019; Cunha e Oliveira, 2020; Oliveira e Miranda, 2020), pesquisas na internet acerca das políticas da rede social estudada, bem como o livro de direito constitucional do autor Bernardo Gonçalves (Gonçalves, 2010). No que diz respeito à teoria de impolidez linguística e o conceito de discurso de ódio teve-se como base teórica o autor Culpeper (2021). Além disso, para tratar do conceito de discurso de ódio foi utilizada a definição elaborada pelo Instagram, bem como as ferramentas utilizadas pela rede social para sua moderação. Na parte empírica, teve-se como base a análise de alguns comentários em postagens do Instagram do jogador Vini Júnior.

Para tanto, o presente texto analisa a relação entre impolidez linguística e discurso de ódio na era digital, destacando a importância de compreender esses fenômenos e entender as dinâmicas comunicativas e sociais online. Inicialmente, na próxima seção, explora-se o conceito de impolidez e sua evolução ao longo do tempo, com destaque para a teoria elaborada por Culpeper. Em seguida, discute-se a interseção entre impolidez e discurso de ódio, enfatizando como comportamentos impolidos podem evoluir para ataques diretos e prejudiciais.

---

<sup>3</sup> Disponível em: [https://www.gov.br/mre/pt-br/canais\\_atendimento/imprensa/notas-a-imprensa/novos-ataques-racistas-na-espanha-a-vinicius-jr](https://www.gov.br/mre/pt-br/canais_atendimento/imprensa/notas-a-imprensa/novos-ataques-racistas-na-espanha-a-vinicius-jr). Acesso em 03 de maio de 2024.

Após, são apresentadas as políticas e padrões formulados pelo Instagram para lidar com o discurso de ódio, incluindo diretrizes claras, sistemas de advertências e a abordagem entre tecnologia e intervenção humana na moderação de conteúdo.

Ato contínuo, analisam-se algumas postagens do jogador Vini Júnior em sua conta no Instagram, avaliando a eficácia desses mecanismos e se, à luz da liberdade de expressão, eles restringem a capacidade dos usuários de se expressarem. Por fim, são apresentadas as considerações finais.

## **2. IMPOLIDEZ LINGUÍSTICA E DISCURSO DE ÓDIO NA ERA DIGITAL**

Conceituar impolidez não é simples, haja vista que sua conceituação permeia, muitas vezes, pela ótica pessoal de quem a define. Essa dificuldade se deve, principalmente, à existência de um grande número de estudos sobre a polidez, com abordagens distintas, e que apesar de demonstrarem a complexidade própria do conceito, prejudicam a compreensão global dessa concepção (Cunha e Oliveira, 2020).

Ainda de acordo com os autores, baseados na proposta de Culpeper (2011), as teorias elaboradas sobre impolidez podem ser divididas em duas fases. A primeira composta pelo aparecimento e solidificação dos estudos desse tema, com estudos feitos por Brown e Levinson (1987), Lakoff (1975; 1977) e Leech (1983). Já na segunda fase, composta por estudos feitos, por exemplo, por Culpeper (2011), Watts (2005) e Terkourafi (2005), procuraram reexaminar os estudos anteriores a fim de expor e resolver as lacunas e problemas teóricos do momento anterior.

Na primeira fase foi dada ênfase ao ato de fala como unidade de análise, muitas vezes de forma descontextualizada, ou seja, sozinha; ao estudo da polidez pautado no locutor e não na interação entre locutor e interlocutor, além de ausência da adoção do contexto no estudo da polidez. Além disso, tradicionalmente, as investigações sobre polidez tinham como foco estratégias destinadas a suavizar a agressividade dos FTAs (atos ameaçadores de fala) em relação às faces positiva e negativa dos interlocutores.

Na segunda fase, foi adotada uma abordagem interacional e discursiva da polidez, superando a ênfase anterior no ato de fala como unidade isolada de análise. Em contraste com a visão tradicional centrada no locutor, os estudiosos dessa fase investigaram a interação entre locutor e interlocutor, bem como a adoção do contexto como elemento fundamental no estudo da polidez. Essa nova perspectiva não apenas inventariou as estratégias de impolidez, mas também explorou suas funções comunicativas em contextos específicos.

Nessa nova abordagem, estudiosos contemporâneos buscam compreender o impacto dos status ou papéis sociais na escolha das estratégias de impolidez, o papel dessas estratégias na definição dos status sociais, a natureza formulaica das estratégias e o papel das instituições sociais na criação e incorporação de estratégias de impolidez. Segundo Cunha e Tomazi (2018), a "virada discursiva" que esse campo tem experimentado é evidenciada pela ampliação de seus interesses e pela incorporação de metodologias e conceitos de áreas afins (*apud* Cunha e Oliveira, 2020).

Com relação à segunda fase, o artigo irá tratar apenas da teoria elaborada por Culpeper. Isso porque a teoria sobre impolidez linguística, conforme apresentada por Culpeper (2011), tem maior compatibilidade conceitual com a definição de discurso de ódio apresentada pelo Instagram, ainda que tais termos tenham significados distintos. Inclusive, sob esse aspecto, vale adiantar que, no caso apresentado neste artigo sobre o jogador Vini Júnior, é possível verificar que o contexto e a intenção, abordados no conceito de impolidez de Culpeper (2011), estão presentes no discurso de ódio proferido a ele.

Nesse sentido, importante diferenciar, de forma resumida, impolidez linguística e discurso de ódio. Segundo Culpeper (2011), a impolidez e o discurso de ódio podem ser abordados por meio das palavras "indelicado" e "odioso", respectivamente. Ambos os termos se diferenciam pelos tipos de emoção aos quais estão associados. O termo odioso, em comparação ao termo indelicado, é caracterizado por comportamentos mais extremos, pela emoção de mágoa, relacionados ao preconceito e um antagonismo em relação a outras pessoas com base no seu pertencimento a um determinado grupo. Já a impolidez se relaciona a comportamentos mais moderados e aos sentimentos de tristeza e raiva.

Segundo Cunha e Oliveira (2020), a impolidez, de acordo com Culpeper (2011), é caracterizada como uma atitude negativa em relação a comportamentos específicos ocorrendo em contextos particulares, revelando-se na interação e, portanto, constitui um fenômeno a ser examinado dentro de um contexto específico, sendo influenciada por expectativas, desejos, crenças sobre a organização social e sua intenção nessa interação. Dessa forma, a avaliação de um comportamento impolido está diretamente ligada à intenção, bem como ao lugar social ocupado pelo locutor em relação aos outros participantes da interação e a diferentes grupos sociais. Importante destacar que, em alguns casos, ainda que o falante não tenha tido a intenção de ser impolido, o ouvinte pode perceber uma intencionalidade de impolidez no ato do locutor.

Para abordar esse fenômeno, Culpeper (2016) propõe uma complementação à teoria da polidez de Brown e Levinson (1987), expandindo as cinco superestratégias de polidez

apresentadas por eles para propor, por sua vez, outras cinco superestratégias de impolidez e uma metaestratégia:

- 1) *Bald on record impoliteness*: o FTA é realizado de uma forma direta, clara, sem ambiguidade e concisa, em circunstâncias em que a preocupação com face não é irrelevante ou minimizada.
- 2) Impolidez positiva: o uso de estratégias que prejudicam os desejos de face positiva do destinatário.
- 3) Impolidez negativa: o uso de estratégias que prejudicam os desejos de face negativa do destinatário.
- 4) Impolidez off-record: o FTA é performado por meio de uma implicatura, mas de tal maneira que uma determinada intenção claramente tem maior peso do que qualquer outra.
- 5) Polidez withhold (retida, negada): não se usam estratégias de polidez em circunstâncias em que esse uso é esperado.

METAESTRATÉGIA DE IMPOLIDEZ Sarcasmo ou falsa (mock) polidez: o FTA é performado com o uso de estratégias de polidez claramente insinceras, que, por isso, se manifestam apenas na superfície do ato. (CUNHA; OLIVEIRA, 2020).

Considerável ainda tecer informações acerca da noção de face utilizada pelo autor. Culpeper adota a noção de face de Spencer-Oatey (2002; 2005), a qual foi definida com base na noção de Goffman e de direitos de sociabilidade.

Segundo Feitosa e Silva (2019), a face, para Goffman (1967), é construída no momento da interação entre os falantes e seus interlocutores, baseando-se nos "atributos sociais aprovados" demonstrados nessa interação. Assim, cada indivíduo procura preservar sua face, buscando atenuar ou manter os atributos socialmente reconhecidos e desejados pelos outros participantes da interação.

Rucero (2014) complementa essa ideia ao explicar que a face é constituída pelos "valores sociais positivos" reivindicados por um ator durante as interações. Essa imagem positiva é mantida por meio de um "trabalho de face", que envolve negociações entre os participantes visando manter os valores atribuídos à face proposta. Essas negociações estão intimamente ligadas ao capital social, já que a legitimação da face durante uma conversação implica o suporte dos demais participantes e o conhecimento das normas de interação.

Assim, a face não é apenas uma representação individual, mas também coletiva, refletindo identidades sociais e culturais. A busca pela manutenção e legitimidade da face influencia diretamente as escolhas discursivas dos envolvidos na interação, moldando os rituais sociais e as dinâmicas comunicativas.

Culpeper (2011) definiu as funções das estratégias que a impolidez pode assumir na interação em três categorias funcionais:

- i) impolidez afetiva: o ofensor expressa de maneira irrestrita suas emoções de raiva, ódio ou cólera em contextos onde essa expressão não é esperada e com o fim de revelar que o alvo de seu ataque é a causa de seu estado emocional;
- ii) impolidez coerciva: o ofensor busca um realinhamento de valores para se beneficiar ou ter seus atuais benefícios reforçados ou protegidos e, por isso, envolve ação coerciva que não é do interesse do alvo do ataque;
- iii) impolidez para entretenimento: o ofensor transforma o interlocutor em alvo de críticas, chacotas e deboches, com o objetivo de divertir um terceiro. De acordo com Culpeper (2011), não há uma relação biunívoca entre formas e funções de impolidez, o que significa que uma mesma estratégia de impolidez, como um xingamento, pode, ao mesmo tempo, ser uma forma de expressar ódio (impolidez afetiva) e de obter poder (impolidez coerciva). (CUNHA, 2019, p.7)

Tem-se como exemplo de impolidez o caso citado em Oliveira e Miranda (2020), o qual oferece um estudo das dinâmicas envolvendo impolidez em um contexto específico de plataforma digital. O artigo concentra-se em um incidente que ocorreu durante uma coletiva de imprensa na qual o então Presidente dos Estados Unidos, Joe Biden, dirigiu-se de forma grosseira ao jornalista Peter Doocy. Tal episódio desencadeou uma série de reações no X (antigo Twitter), que se tornou um espaço para debates acerca da impolidez e da linguagem vergonhosa na arena política.

Tal artigo analisou diversos comentários proferidos pelos usuários e constatou que os mesmos foram majoritariamente impolidos. Essas manifestações podem ser categorizadas dentro das funções de estratégia afetiva ou coercitiva elaboradas por Culpeper (2011). Os que expressam raiva, desgosto ou desprezo em relação ao jornalista ou ao presidente se encaixariam na categoria de impolidez afetiva. Por outro lado, os que buscam reforçar o poder do presidente, desencorajar críticas futuras ou manipular a opinião pública em favor do governo podem ser classificados como impolidez coercitiva. Essas estratégias impolidas têm o objetivo de transmitir emoções negativas ou pressionar o receptor a agir de acordo com os interesses do ofensor, demonstrando assim as complexas dinâmicas envolvidas na interação online e na arena política.

A interseção entre impolidez e discurso de ódio revela-se fundamental para compreender as dinâmicas sociais e comunicativas, especialmente em ambientes digitais. A impolidez, muitas vezes, serve como precursora ou facilitadora do discurso de ódio, uma vez que comportamentos considerados impolidos podem evoluir para ataques diretos e hostis com base em características protegidas.

Embora não tenha sido verificada a presença de discurso de ódio nos tweets do mencionado artigo, percebe-se que um espaço digital extremamente impolido contribui para a criação de um ambiente hostil onde ideias extremistas e prejudiciais podem prosperar. Esses

ambientes digitais hostis tornam-se propícios para o surgimento e a disseminação do discurso de ódio, pois os comentários impolidos alimentam a polarização e o antagonismo, incentivando ataques verbais e comportamentos agressivos. Dessa forma, é essencial reconhecer e combater a impolidez online como parte do esforço mais amplo para prevenir e mitigar o discurso de ódio nas plataformas digitais.

Paralelamente, a compreensão das políticas e definições de discurso de ódio, como as estabelecidas pelo Instagram, lança luz sobre as manifestações digitais da impolidez, que, inclusive, são confundidas com o discurso de ódio, em razão da popularidade que esse termo alcançou, exacerbando a natureza prejudicial e ofensiva dessas interações. Assim, a análise conjunta desses fenômenos permite uma compreensão mais completa das dinâmicas comunicativas e sociais na era digital.

O Instagram, em suas Políticas e Padrões da Comunidade da rede, acessíveis em seu website, delimita o discurso de ódio como:

um ataque direto a pessoas, e não a conceitos e instituições, baseado no que chamamos de características protegidas: raça, etnia, nacionalidade, deficiência, religião, casta, orientação sexual, sexo, identidade de gênero e doença grave. Definimos ataques como discursos violentos ou desumanizantes, estereótipos prejudiciais, declarações de inferioridade, expressões de desprezo, repulsa ou rejeição, xingamentos e incitações à exclusão ou segregação. Também proibimos o uso de estereótipos prejudiciais, que definimos como comparações desumanizantes historicamente usadas para atacar, intimidar ou excluir grupos específicos, e que muitas vezes estão ligadas à violência no meio físico. Consideramos a idade uma característica protegida quando referenciada juntamente com outra característica também protegida. Também protegemos de ataques mais graves refugiados, migrantes, imigrantes e pessoas que buscam asilo, embora permitamos comentários e críticas às políticas de imigração. Da mesma forma, fornecemos algumas proteções para aspectos como ocupação, quando eles são mencionados juntamente com uma característica protegida. Às vezes, com base em particularidades locais, consideramos palavras ou frases específicas como proxies usados com frequência para grupos com características protegidas. Além disso, proibimos o uso de calúnias usadas para atacar pessoas com base em características protegidas.<sup>4</sup>

Conforme delineado por Culpeper (2011), impolidez e discurso de ódio são termos com significados diferentes. O primeiro pode ser associado à palavra "indelicado" e o segundo à palavra "odioso". Enquanto a impolidez está ligada a comportamentos moderados e sentimentos de tristeza e raiva, o discurso de ódio envolve comportamentos extremos, motivados por emoções de mágoa, preconceito e antagonismo em relação a indivíduos com base em seu pertencimento a determinados grupos. O Instagram, em suas Políticas e Padrões da Comunidade, adota uma abordagem similar, delimitando o discurso de ódio como ataques

---

<sup>4</sup> Disponível em: <https://transparency.meta.com/pt-br/policies/community-standards/hate-speech/>. Acesso em 23 de abril de 2024.

diretos a pessoas com base em características protegidas, como raça, etnia, religião, orientação sexual, entre outras. Assim como Culpeper (2011) diferencia a impolidez do discurso de ódio, o Instagram distingue comentários impolidos de ataques diretos baseados em atributos tutelados, estabelecendo um paralelo entre os dois conceitos. Essa abordagem reflete a preocupação em distinguir entre expressões impolidas e discursos que visam prejudicar ou desumanizar indivíduos com base em particularidades resguardadas, alinhando-se com as definições propostas por Culpeper (2011).

Importante destacar que o discurso de ódio infringe e permeia os limites da liberdade de expressão, sendo necessário moderá-lo devido às violações aos direitos humanos que representa. Diferentemente da impolidez, que está associada a comportamentos mais brandos e permeia a liberdade de expressão, o discurso de ódio está sempre ligado a um atentado aos Direitos Humanos, exigindo medidas regulatórias e moderadoras, haja vista que essa forma de expressão configura-se como crime. Uma abordagem sugerida para diferenciar casos de impolidez linguística dos casos de discurso de ódio, especialmente em situações mais complexas de discernimento, é a consideração das definições propostas por Culpeper (2011). É importante observar que, na análise desses fenômenos, a intencionalidade desempenha um papel crucial, sendo fundamental para a aplicação das normas sociais.

### **3. MEDIDAS DE MODERAÇÃO DO INSTAGRAM CONTRA O DISCURSO DE ÓDIO**

Neste tópico, serão examinados os meios implementados pelo Instagram para combater o discurso de ódio em sua plataforma. Serão detalhadas as diretrizes e políticas adotadas pela rede social, assim como as tecnologias e abordagens utilizadas para identificar e mitigar conteúdos prejudiciais. Ainda, será feita uma descrição de como essas medidas operam na prática e sua eficácia na promoção de um ambiente online mais seguro e respeitoso.

As diretrizes e políticas do Instagram que utilizamos neste trabalho estão disponíveis na página "Terms and Policies" e em "Transparency Center". Os usuários podem acessá-las diretamente através do link: <https://transparency.fb.com/pt-br/>. Para uma análise mais específica das diretrizes do Instagram, basta clicar na opção "Políticas" na parte superior da página e selecionar "Diretrizes da Comunidade do Instagram". Além disso, na mesma página, na seção superior, está disponível a opção "Verificação", que fornece informações sobre a detecção de violações e as medidas aplicadas pelo Instagram em resposta a tais violações.

As Diretrizes da Comunidade do Instagram abordam uma variedade de questões importantes para manter um ambiente seguro e respeitoso na plataforma. Durante a pandemia de COVID-19, o Instagram implementou medidas específicas para proteger os usuários de conteúdo prejudicial relacionado à crise de saúde global. Isso inclui políticas que proíbem a disseminação de desinformação sobre a pandemia, bem como a coordenação de danos que possam resultar em consequências reais. Além disso, o Instagram trabalha para remover conteúdo que viole seus padrões, incluindo discurso de ódio, bullying e assédio, assim como combater a venda ilegal de itens relacionados à proteção contra o vírus.<sup>5</sup>

As diretrizes também tratam de outras questões relevantes, como a publicação de conteúdo autêntico e respeito às leis locais e internacionais. Há restrições claras quanto à nudez, com exceções limitadas para situações como arte ou conscientização sobre questões de saúde. O Instagram proíbe explicitamente em suas diretrizes e políticas a promoção da violência, o discurso de ódio e a exploração de vulnerabilidades, na tentativa de criar um ambiente online positivo e inclusivo.<sup>6</sup>

A plataforma incentiva a colaboração entre a comunidade e a equipe do Instagram, encorajando os usuários a relatarem qualquer conteúdo que viole as diretrizes. O compromisso ao qual eles se propõem é o de responder rapidamente e remover o conteúdo inadequado. Essas diretrizes refletem não apenas um esforço para manter a segurança dos usuários, mas também para promover uma cultura de respeito e responsabilidade online.<sup>7</sup>

O Instagram, como parte da empresa Meta, afirma implementar políticas rigorosas para detectar e lidar com violações de conteúdo em sua plataforma. Os Padrões da Comunidade do Instagram estabelecem diretrizes claras sobre o que é permitido e o que não é permitido na rede social. Quando o conteúdo violador é identificado, a Meta propõe-se a removê-lo assim que detectado, seja por denúncia de usuários ou por sistemas automatizados.<sup>8</sup>

Um sistema de advertências é empregado para monitorar e responsabilizar os usuários por violações de conteúdo. Dependendo da gravidade e recorrência das violações, podem ser aplicadas restrições à conta, como a perda de acesso a determinados recursos ou até mesmo a

---

<sup>5</sup> Disponível em: <https://www.facebook.com/help/instagram/477434105621119/>. Acesso em: 03 de maio de 2024.

<sup>6</sup> Disponível em: <https://www.facebook.com/help/instagram/477434105621119/>. Acesso em: 03 de maio de 2024.

<sup>7</sup> Disponível em: <https://www.facebook.com/help/instagram/477434105621119/>. Acesso em: 03 de maio de 2024

<sup>8</sup> Disponível em: <https://transparency.meta.com/pt-br/policies/>. Acesso em: 03 de maio de 2024.

desativação da conta. Essas restrições são progressivas, começando com um aviso para a primeira violação e escalando para restrições mais severas em caso de reincidência.<sup>9</sup>

Quando o conteúdo é removido, o usuário é notificado sobre a violação e o motivo da remoção, além de receber orientações sobre como evitar futuras violações. A Meta diz se esforçar para fornecer detalhes sobre qual parte das políticas foi violada, de modo a ajudar os usuários a entenderem melhor as diretrizes da comunidade.<sup>10</sup>

Além disso, a empresa assegura que realiza análises detalhadas sobre o uso de calúnias e palavras ofensivas nas plataformas, utilizando equipes regionais e especialistas em políticas para determinar quais termos devem ser restringidos com base em considerações culturais e linguísticas. O anúncio é o de que essas análises sejam constantemente atualizadas e refinadas para garantir que as listas de calúnias sejam abrangentes e atualizadas.<sup>11</sup>

Segundo eles, a aplicação das políticas é realizada por meio de uma combinação de tecnologia e equipes de análise humana. A tecnologia se propõe a uma detecção proativa de conteúdo violador, enquanto as equipes compostas por seres humanos realizam análises mais detalhadas e tomam decisões finais sobre a remoção de conteúdo.<sup>12</sup>

Outrossim, os usuários podem configurar seus perfis como privados, restringir temporariamente comentários e mensagens, ocultar palavras ofensivas, bloquear comentários de determinados usuários e limitar menções em suas postagens. Além disso, podem optar por ter contas restritas, bloqueadas ou silenciadas, permitindo aos usuários controlarem sua interação com outros perfis e oferecendo diferentes níveis de privacidade e proteção contra assédio online.<sup>13</sup>

Em resumo, conforme explicitado em seus termos, o Instagram adota medidas abrangentes para garantir um ambiente online seguro e saudável para seus usuários, incluindo políticas claras, sistemas de advertências, análises detalhadas e uma abordagem equilibrada entre tecnologia e intervenção humana na moderação de conteúdo.

---

<sup>9</sup> Disponível em: <https://transparency.meta.com/pt-br/policies/community-standards/hate-speech/>. Acesso em 03 de maio de 2024.

<sup>10</sup> Disponível em: <https://transparency.meta.com/pt-br/enforcement/>. Acesso em: 03 de maio de 2024.

<sup>11</sup> Disponível em: <https://transparency.meta.com/pt-br/enforcement/detecting-violations/challenge-of-detecting-hate-speech>. Acesso em: 03 de maio de 2024.

<sup>12</sup> Disponível e: <https://transparency.meta.com/pt-br/enforcement/detecting-violations/how-review-teams-work/>. Acesso em: 03 de maio de 2024.

<sup>13</sup> Disponível em: <https://www.facebook.com/help/instagram/477434105621119/>. Acesso em: 03 de maio de 2024.

#### **4. ANÁLISE DAS MEDIDAS DO INSTAGRAM: MODERAÇÃO E LIBERDADE DE EXPRESSÃO**

De acordo com Caldeira (2022), as ações moderadoras implementadas pelo Instagram contribuíram para a diminuição das ofensas e dos discursos de ódio na plataforma. Entretanto, a nível legislativo, algumas falhas persistem, como evidenciado pelo Marco Civil da Internet. Nem sempre as medidas de moderação conseguem barrar todo o conteúdo prejudicial, o que frequentemente resulta em disputas judiciais. A importância do Marco Civil da Internet no contexto brasileiro se destaca como uma base legal essencial para lidar com essas questões, apesar das dificuldades em sua implementação e eficácia.

Segundo Lottenberg e Vainzof (2018), após a implementação do Marco Civil da Internet, observou-se uma mudança significativa na jurisprudência brasileira em relação à responsabilidade dos provedores de aplicação sobre conteúdos ofensivos. Antes disso, a legislação exigia que os provedores tomassem medidas rápidas para coibir a disseminação de conteúdos depreciativos, suspendendo preventivamente as respectivas páginas até uma análise mais aprofundada. No entanto, após a vigência do Marco Civil, a responsabilidade dos provedores passou a ser postergada até a notificação judicial específica para remoção do conteúdo. Essa mudança reflete a preocupação em conciliar a liberdade de expressão com a proteção dos direitos individuais, estabelecendo um processo mais formal e jurídico para resolver conflitos online.

No entanto, a legislação trouxe novos desafios e críticas, especialmente em relação ao seu impacto na proteção dos direitos dos usuários da Internet. Lottenberg e Vainzof (2018) argumentam que o Marco Civil estabeleceu um mecanismo engessado que favorece excessivamente as empresas que exploram redes sociais, em detrimento dos direitos dos usuários. Além disso, a legislação parece priorizar a liberdade de expressão sobre outros direitos fundamentais, como a honra e a imagem das pessoas, gerando insegurança jurídica e questionamentos sobre sua constitucionalidade.

É importante destacar que o Marco Civil da Internet também influenciou a atuação dos provedores de aplicação, que agora são incentivados a adotar medidas internas para resolver extrajudicialmente disputas sobre conteúdos ilícitos. Essa abordagem busca conciliar a necessidade de combater o discurso de ódio e outras formas de conteúdo prejudicial com a promoção da liberdade de expressão e o respeito aos direitos individuais. No entanto, ainda há questões a serem resolvidas, como a eficácia dessas medidas na prática e a necessidade de garantir uma proteção adequada aos usuários da Internet.

Para analisar as medidas do Instagram destinadas a limitar o discurso de ódio e a impolidez, é imprescindível contextualizá-las dentro do conceito de liberdade de expressão. De acordo com o professor Bernardo Gonçalves (2010), em seu Manual de Direito Constitucional, a liberdade de expressão é considerada um dos pilares fundamentais da democracia, conferindo aos indivíduos o direito de expressar suas opiniões e ideias sem interferência do Estado ou de terceiros. Esse direito, consagrado em diversas constituições ao redor do mundo, é essencial para o livre debate de ideias e para o funcionamento saudável das sociedades democráticas.

No entanto, a liberdade de expressão não é absoluta e deve ser ponderada com outros princípios e valores igualmente relevantes. Juristas norte-americanos tendem a defender uma interpretação ampla desse direito, sustentando que ele deve ser preservado mesmo diante de discursos considerados ofensivos ou desagradáveis. Por outro lado, juristas europeus são mais propensos a aceitar restrições ao discurso de ódio, baseando-se na convicção de que tais restrições podem promover a igualdade e proteger a dignidade humana. Essa diferença de abordagem reflete diferentes entendimentos sobre o equilíbrio entre a liberdade individual e o interesse coletivo na manutenção de uma sociedade justa e inclusiva.

Dentro desse contexto, as medidas do Instagram que visam limitar o discurso de ódio e a impolidez representam uma tentativa da plataforma de equilibrar a liberdade de expressão com a proteção dos usuários contra conteúdos prejudiciais. Por exemplo, a opção de configurar o perfil como privado permite aos usuários controlar quem pode visualizar seu conteúdo, proporcionando uma camada adicional de privacidade e segurança. Além disso, a capacidade de filtrar comentários e mensagens indesejadas oferece aos usuários maior controle sobre as interações que desejam receber, reduzindo assim a exposição a conteúdos ofensivos ou agressivos.

No entanto, é importante ressaltar que as medidas do Instagram não são capazes de resolver completamente o problema do discurso de ódio e da impolidez na plataforma. Um exemplo claro disso é observado em perfis de pessoas famosas, como o jogador de futebol Vini Júnior, do time espanhol Real Madrid, que é alvo de uma quantidade significativa de comentários racistas e ofensivos, apesar das políticas de moderação implementadas pela rede social.

Cabe mencionar aqui a quantidade alarmante de emojis de macaco, gorila e cocô que uma simples postagem, contendo a foto do jogador, atrai. Essa observação evidencia a ausência de moderação perante uma volumosa quantidade de comentários, notoriamente repetitivos e,

muito provavelmente, gerados por contas automatizadas. Detalhar esses ataques revela a natureza intrinsecamente discriminatória e hostil dos mesmos. As agressões diretas à cor de pele do jogador, como “aquí esta tu banana” (Figura 1), “negrito” (Figura 2), “mono macaco de merda” (Figura 3), e a proliferação de emojis de macaco (Figura 4), manifestam uma flagrante violação das políticas antidiscriminação da plataforma, como pode ser visto nos seguintes prints retirados do próprio Instagram do jogador:

**Figura 1 – Print de publicação no Instagram do jogador Vinicius Junior**



Fonte: INSTAGRAM – Print de publicação no Instagram do Jogador Vinicius Junior. Disponível em: <https://www.instagram.com/vinijr/>. Acesso em: 10 de maio de 2024.

**Figura 2 – Print de publicação no Instagram do jogador Vinicius Junior**



Fonte: INSTAGRAM – Print de publicação no Instagram do Jogador Vinicius Junior. Disponível em: <https://www.instagram.com/vinijr/>. Acesso em: 10 de maio de 2024.

**Figura 3 – Print de publicação no Instagram do jogador Vinicius Junior**



Fonte: INSTAGRAM – Print de publicação no Instagram do Jogador Vinicius Junior. Disponível em: <https://www.instagram.com/vinijr/>. Acesso em: 10 de maio de 2024.

**Figura 4 – Print de publicação no Instagram do jogador Vinicius Junior**



Fonte: INSTAGRAM – Print de publicação no Instagram do Jogador Vinicius Junior. Disponível em: <https://www.instagram.com/vinijr/>. Acesso em: 10 de maio de 2024.

Tem-se como exemplos de superestratégia "Bald on record impoliteness", os ataques “aquí esta tu banana”, “negrito” e “mono macaco de merda”, uma vez que as agressões mencionadas são realizada de maneira direta, clara e sem qualquer tentativa de disfarçar a intenção ofensiva, característica dessa superestratégia. Estes exemplos<sup>14</sup> são emblemáticos do discurso de ódio, uma vez que atacam características protegidas, tais como raça e etnia, conforme definido por Culpeper (2011). Diferenciando-se da impolidez, que pode implicar em comportamentos menos extremos, o discurso de ódio é caracterizado por uma intenção explícita de hostilidade e um antagonismo manifesto. Essa distinção é de suma importância para o entendimento e ações eficazes no combate a essas formas de violência online, ressaltando a urgência de uma moderação mais eficaz por parte das plataformas de mídia social e, em especial, por parte do Instagram.

É desconcertante perceber que, mesmo em um perfil de grande relevância, com 46,1 milhões de seguidores, a moderação do Instagram parece falhar em detectar e conter a enorme quantidade de ódio propagada, apesar de todas as medidas a que a plataforma de propõe a implementar para proteger os usuários e banir esse tipo de discurso. Este é apenas um exemplo dentre muitos que destacam os desafios enfrentados pelas plataformas online na contenção do discurso de ódio e na promoção de um ambiente seguro e inclusivo para todos os usuários.

A eficácia das medidas do Instagram na moderação do discurso de ódio e da impolidez depende não apenas de sua implementação técnica, mas também de uma abordagem holística que leve em consideração os diferentes aspectos envolvidos. Isso inclui a necessidade de promover uma cultura de respeito e tolerância entre os usuários, bem como o desenvolvimento contínuo de políticas e tecnologias de moderação mais sofisticadas

<sup>14</sup> Os prints completos podem ser encontrados nos Anexos I a IX.

Ao mesmo tempo, é crucial manter um equilíbrio adequado entre a liberdade de expressão e a proteção dos direitos e da dignidade dos indivíduos, garantindo assim uma experiência online segura e inclusiva para todos os usuários. Isso pode ser alcançado através de uma abordagem multifacetada que envolve diversos aspectos. Nesse sentido, a educação e conscientização são fundamentais para promover uma cultura de respeito e empatia online, capacitando os usuários a reconhecerem e responderem adequadamente ao discurso de ódio. Além disso, a implementação eficaz de políticas e normas de comunidade claras e rigorosas, que proíbam explicitamente o discurso de ódio, é essencial para estabelecer limites claros e promover um ambiente online saudável.

O uso de ferramentas de moderação avançadas, como inteligência artificial e aprendizado de máquina, pode ajudar as plataformas a identificar e remover conteúdos prejudiciais de forma mais ágil e eficiente. A responsabilidade das plataformas também desempenha um papel crucial, pois elas devem assumir a responsabilidade por criar e manter ambientes online seguros, investindo em recursos e tecnologias para combater o discurso de ódio.

Além disso, as parcerias com organizações da sociedade civil, como grupos de defesa dos direitos humanos ou especialistas em segurança cibernética, podem fornecer ferramentas valiosas e apoio na elaboração e implementação de estratégias eficazes de combate ao discurso de ódio. Por fim, é essencial fornecer apoio adequado às vítimas de discurso de ódio, oferecendo recursos de suporte emocional, orientação legal e assistência para lidar com as consequências psicológicas e sociais desse tipo de violência online. Essas medidas combinadas podem contribuir significativamente para criar um ambiente online mais seguro, inclusivo e respeitoso para todos.

## **5. CONSIDERAÇÕES FINAIS**

A evolução da internet trouxe consigo uma série de desafios, e o surgimento das redes sociais não é exceção. O Instagram, em particular, emergiu como um palco virtual para diversas formas de expressão, incluindo o discurso de ódio. Enquanto as redes sociais oferecem um espaço valioso para a interação e conexão positiva entre os usuários, também se tornaram ambientes propícios para a disseminação de discursos prejudiciais e ofensivos.

Tal realidade provocou um desafio delicado a ser enfrentado pelo Instagram: moderar o discurso de ódio sem cercear a liberdade de expressão dos usuários. Embora seja crucial

proteger os usuários contra abusos e violência online, também é necessário garantir que as plataformas não limitem indevidamente a livre troca de ideias e opiniões legítimas.

Diante desse novo cenário foi proposto este estudo tendo sido levantada a seguinte questão: Há moderação eficaz do discurso de ódio em comentários nas postagens do Instagram? Teve-se como hipótese deste estudo que os mecanismos de moderação implementados pelo Instagram não conseguem efetivamente mitigar a presença de discurso de ódio em suas postagens.

Para responder a questão levantada, foi feita uma análise da relação entre impolidez linguística e discurso de ódio na era digital, destacando a importância de compreender esses fenômenos e entender as dinâmicas comunicativas e sociais online. Explorou-se o conceito de impolidez e sua evolução ao longo do tempo, com destaque para a teoria elaborada por Culpeper (2021). Em seguida, discutiu-se sobre a interseção entre impolidez e discurso de ódio, enfatizando como comportamentos impolidos podem evoluir para ataques diretos e prejudiciais. Após, foram apresentadas as políticas e padrões formulados pelo Instagram para lidar com o discurso de ódio, incluindo diretrizes claras, sistemas de advertências e a abordagem entre tecnologia e intervenção humana na moderação de conteúdo. Ato contínuo, foram analisadas algumas postagens do jogador Vini Júnior em sua conta no Instagram, avaliando a eficácia desses mecanismos e se, à luz da liberdade de expressão, eles restringem a capacidade dos usuários de se expressarem.

A análise das diretrizes e políticas do Instagram revelou um conjunto abrangente de medidas implementadas para combater o discurso de ódio e promover um ambiente online seguro e respeitoso. No entanto, apesar desses esforços, persistem desafios significativos, como evidenciado pelo contínuo fluxo de conteúdo prejudicial.

A eficácia das medidas do Instagram na moderação do discurso de ódio depende não apenas de sua implementação técnica, mas também de uma abordagem holística que leve em consideração diversos fatores. Isso inclui a promoção de uma cultura de respeito e tolerância entre os usuários, o desenvolvimento contínuo de políticas e tecnologias de moderação mais sofisticadas e o equilíbrio adequado entre a liberdade de expressão e a proteção dos direitos e da dignidade dos indivíduos.

No contexto das discussões sobre a moderação de conteúdo online, é essencial incorporar as definições propostas por Culpeper para distinguir o conceito de impolidez linguística, que se enquadra no domínio da liberdade de expressão, do conceito de discurso de

ódio, caracterizado como um crime e um atentado aos Direitos Humanos. Essa distinção se torna particularmente relevante em situações que envolvem comentários que se situam em uma "área cinzenta", onde é difícil discernir se constituem discurso de ódio ou simplesmente expressões polêmicas. Por meio das definições estabelecidas por Culpeper, é possível diferenciar esses conceitos e tomar decisões mais embasadas em relação à moderação de conteúdo.

Portanto, ao analisar casos em que há ambiguidade quanto à natureza dos comentários, é fundamental considerar cuidadosamente se estão mais próximos da impolidez linguística, caracterizada por violações menos graves e que podem ser toleradas em um contexto de liberdade de expressão, ou se configuram efetivamente como discurso de ódio, demandando uma ação mais enérgica para proteger os direitos e a dignidade dos usuários afetados. A aplicação dessas definições permite uma abordagem mais precisa e justa na moderação de conteúdo, garantindo ao mesmo tempo a promoção de um ambiente online mais seguro e inclusivo.

Assim, o desafio de garantir uma moderação eficaz do discurso de ódio no Instagram e em outras plataformas de mídia social requer uma abordagem colaborativa e multifacetada, envolvendo não apenas as empresas responsáveis, mas também os usuários, legisladores e a sociedade em geral. Somente por meio de esforços conjuntos e contínuos pode-se criar um ambiente online verdadeiramente inclusivo e respeitoso para todos os usuários.

## REFERÊNCIAS

ARAÚJO, M. F. Débora. **Estratégias de (im)polidez em comentários de notícias publicadas no Facebook**. Dissertação de mestrado. Universidade Federal de Minas Gerais, Faculdade de Letras. 128 páginas. 2022. Acesso em: 15 dez. 2023.

BROWN, P; LEVINSON, S. **Politeness: some universals in language use**. Cambridge: Cambridge University Press, 1987.

CALDEIRA, Q. Íris. **Discursos de ódio no Instagram: Estudo-caso Carolina Deslandes**. Dissertação de mestrado. Instituto Universitário de Lisboa. 2022. Acesso em: 15 dez. 2023.

CESPEDES, Fernando et al. **Intolerâncias nas redes: um problema crescente**. Conexão Pública. Disponível em: <https://conexaopublica.com.br/intolerancias-nas-redes-um-problema-crescente/>. Acesso em: 03 de maio de 2024.

CULPEPER, J. **Impoliteness: using language to cause offense**. Cambridge: Cambridge University Press, 2011.

CUNHA, G. X.; OLIVEIRA, A. L. A. M. **Teorias de im/polidez linguística: revisitando o estado da arte para uma contribuição teórica sobre o tema (Theories of linguistic im/politeness: revisiting the state of the art for a theoretical contribution on the topic)**. Estudos da Língua(gem), [S. l.], v. 18, n. 2, p. 135-162, 2020. DOI: 10.22481/el.v18i2.6409. Disponível em: <https://periodicos2.uesb.br/index.php/estudosdalinguagem/article/view/6409>. Acesso em: 12 mar. 2024.

CUNHA, G. X.; TOMAZI, M. M. **O uso agressivo da linguagem em uma audiência: uma abordagem discursiva e interacionista para o estudo da im/polidez**. Calidoscópio, Porto Alegre, v. 17.2, p. 297-319, 2018.

Facebook. **Ajuda do Instagram: Padrões da Comunidade**. Disponível em: <https://www.facebook.com/help/instagram/477434105621119/>. Acesso em: 03 de maio de 2024.

FEITOSA, G. M., & SILVA, T. M. O. (2019). **Do preconceito à (im)polidez: Aspectos sociais, ideológicos e linguísticos que circunscrevem práticas racistas e sexistas no Facebook**. Cadernos de Linguagem e Sociedade, 20(especial).

FERNADES, Bernardo Gonçalves. **Curso de Direito Constitucional**. 3. ed. São Paulo: Saraiva, 2010.

GOFFMAN, E. **La mise en scène de la vie quotidienne: les relations em public**. v. 2. Paris: Les Éditions de Minuit, 1973.

GOFFMAN, E. On face-work: an analysis of ritual elements in social interaction. In: GOFFMAN, E. **Interaction Ritual: essays on face-to-face behavior**. New York: Pantheon Books, 1967[1955]. p. 5-45.

KERBRAT-ORECCHIONI, C. **L'impolitesse en interaction : aperçus théoriques et étude de cas**. Journal in English Lexicology, v. 2, p. 35-60, 2010.

KERBRAT-ORECCHIONI, C. **Politeness, impoliteness, non-politeness, “polirudeness”: the case of political TV debates**. In: JAMET, D.; JOBERT, M. (Orgs.) Aspects of linguistic impoliteness. Cambridge: Cambridge Scholars Publishing. 2013. p. 16-45.

LAKOFF, R. **Language and woman's place**. New York: Harper Colophon Books, 1975.

LAKOFF, R. What you can do with words: politeness, pragmatics and performatives. In: ROGERS, A.; WALL, B; MURPHY, J. P. (Org.) **Proceedings of the Texas Conference on performatives, presuppositions and implicatues**. Arlington: Center for Applied Linguistics, p. 94-120, 1977.

LEECH, G. **Principles of Pragmatics**. London: Longman, 1983.

LOTTENBERG, M. Opinião: **Discurso de Ódio nas Redes Sociais e o Marco Civil da Internet (Parte 2)**. Conjur, 13 Jul. 2018. Disponível em: <https://www.conjur.com.br/2018-jul-13/opiniaio-discurso-odio-redes-sociais-marco-civil-parte2/>. Acesso em: 03 maio 2024.

Meta. **Aplicação das Políticas.** Disponível em: <https://transparency.meta.com/pt-br/enforcement/>. Acesso em: 03 de maio de 2024.

Meta. **Como as Equipes de Revisão Funcionam.** Disponível em: <https://transparency.meta.com/pt-br/enforcement/detecting-violations/how-review-teams-work/>. Acesso em: 03 de maio de 2024.

Meta. **Desafios na Detecção de Violações.** Disponível em: <https://transparency.meta.com/pt-br/enforcement/detecting-violations/challenge-of-detecting-hate-speech/>. Acesso em: 03 de maio de 2024.

Meta. **Políticas de Comunidade: Discurso de Ódio.** Disponível em: <https://transparency.meta.com/pt-br/policies/community-standards/hate-speech/>. Acesso em: 03 de maio de 2024.

Meta. **Políticas do Meta.** Disponível em: <https://transparency.meta.com/pt-br/policies/>. Acesso em: 03 de maio de 2024.

Ministério das Relações Exteriores. **Novos ataques racistas na Espanha a Vinicius Jr.** Disponível em: [https://www.gov.br/mre/pt-br/canais\\_atendimento/imprensa/notas-a-imprensa/novos-ataques-racistas-na-espanha-a-vinicius-jr](https://www.gov.br/mre/pt-br/canais_atendimento/imprensa/notas-a-imprensa/novos-ataques-racistas-na-espanha-a-vinicius-jr). Acesso em: 03 de maio de 2024.

NAHUZ, F. **World Wide Web: Aspectos Teóricos Dos Mecanismos De Busca.** Informação & Sociedade, [S. l.], v. 9, n. 2, 1999. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/384>. Acesso em: 20 mar. 2024.

OLIVEIRA, A. L. A. M., & Miranda, M. V. (2022). **“Calling a spade, a spade”:** **Impoliteness and shame on Twitter.** Journal of Research in Applied Linguistics, 13(2), 22-32. <https://doi.org/10.22055/RALS.2022.17800>. Acesso em: 12 mar. 2024.

RUCUERO, R. (2014). **Curtir, compartilhar, comentar: trabalho de face, conversação e redes sociais no Facebook.** Verso e Reverso, 28(68), maio-agosto 2014, 114-124.

SANTOS, Felipe. **Conversion.** Conversion. Disponível em: <https://www.conversion.com.br/blog/redes-sociais/>. Acesso em: 03 de maio de 2024.

SPENCER-OATEY, H. Managing rapport in talk: using rapport sensitive incidents to explore the motivational concerns underlying the management of relations. Journal of Pragmatics, v. 34, p. 529-545, 2002.

SPENCER-OATEY, H. **(Im)politeness, face and perceptions of rapport: unpackaging their bases and interrelationships.** Journal of politeness research, v. 1, p. 95-119, 2005.

TERKOURAFI, M. **Beyond the micro-level in politeness research.** Journal of politeness research, v. 1, p. 237-262, 2005.

WATTS, R. J.. Linguistic politeness research: Quo vadis? In: WATTS, R. J.; SACHIKO, I.; EHLICH, K. (Org.) **Politeness in Language: studies in its History,** Theory and Practice. Berlin: Mouton de Gruyter, 2005. p. 11-47.

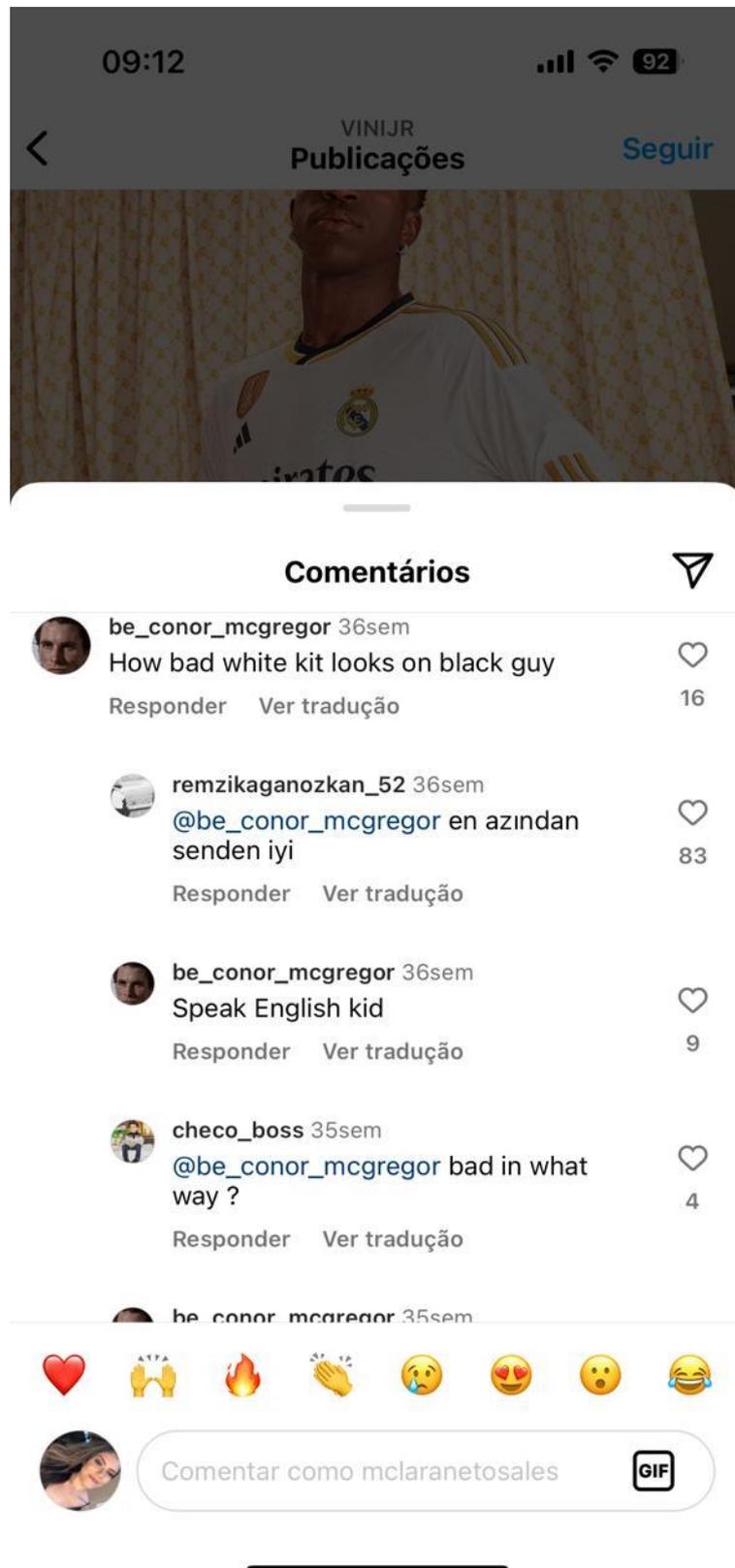
## 6. ANEXOS

### ANEXO I – PRINT DE PUBLICAÇÃO NO INSTAGRAM DO JOGADOR VINICIUS JUNIOR



Fonte: INSTAGRAM – Print de publicação no Instagram do Jogador Vinicius Junior.  
Disponível em: <https://www.instagram.com/vinijr/>. Acesso em: 10 de maio de 2024.

## ANEXO II – PRINT DE PUBLICAÇÃO NO INSTAGRAM DO JOGADOR VINICIUS JUNIOR



Fonte: INSTAGRAM – Print de publicação no Instagram do Jogador Vinicius Junior.  
Disponível em: <https://www.instagram.com/vinijr/>. Acesso em: 10 de maio de 2024.

### ANEXO III – PRINT DE PUBLICAÇÃO NO INSTAGRAM DO JOGADOR VINICIUS JUNIOR



Fonte: INSTAGRAM – Print de publicação no Instagram do Jogador Vinicius Junior.  
Disponível em: <https://www.instagram.com/vinijr/>. Acesso em: 10 de maio de 2024.

## ANEXO IV – PRINT DE PUBLICAÇÃO NO INSTAGRAM DO JOGADOR VINICIUS JUNIOR



Fonte: INSTAGRAM – Print de publicação no Instagram do Jogador Vinicius Junior. Disponível em: <https://www.instagram.com/vinijr/>. Acesso em: 10 de maio de 2024.

### ANEXO V – PRINT DE PUBLICAÇÃO NO INSTAGRAM DO JOGADOR VINICIUS JUNIOR



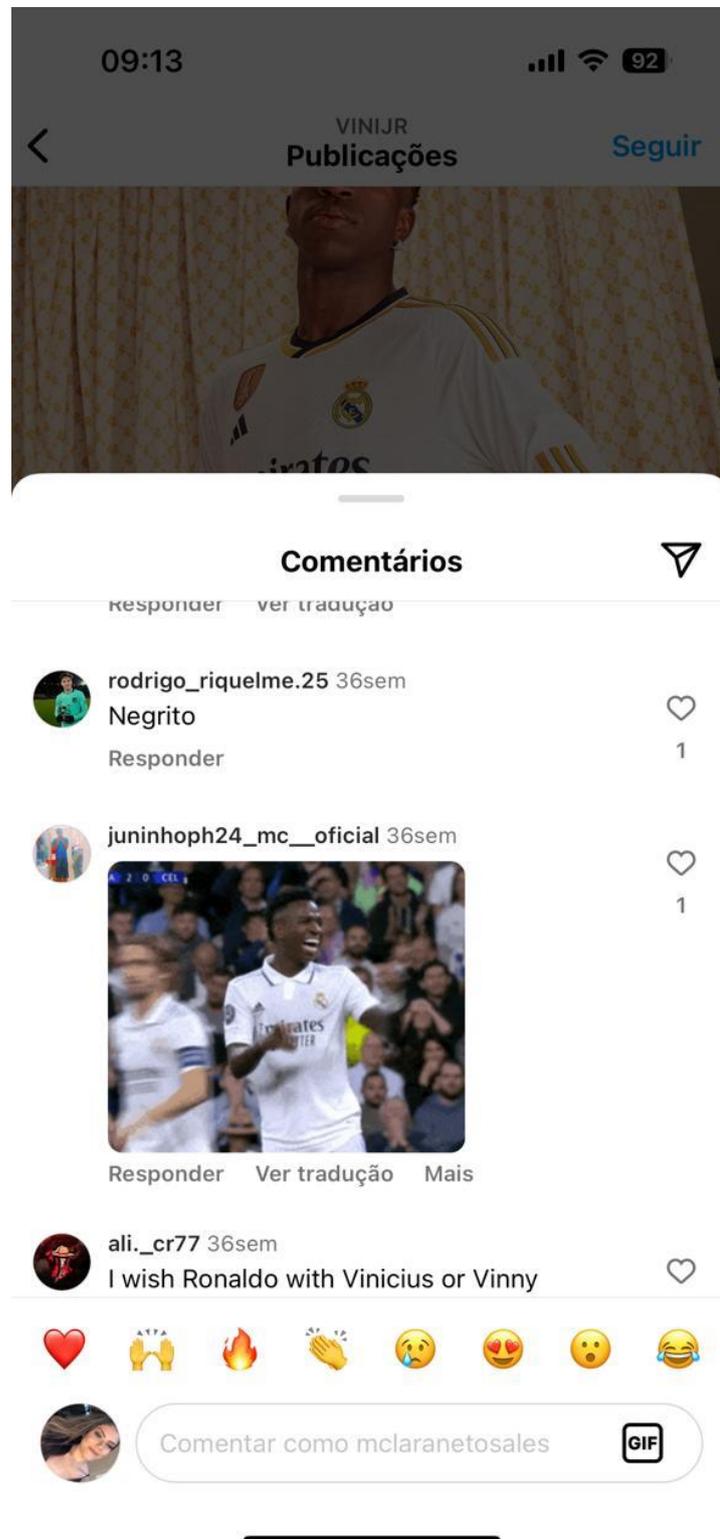
Fonte: INSTAGRAM – Print de publicação no Instagram do Jogador Vinicius Junior. Disponível em: <https://www.instagram.com/vinijr/>. Acesso em: 10 de maio de 2024.

## ANEXO VI – PRINT DE PUBLICAÇÃO NO INSTAGRAM DO JOGADOR VINICIUS JUNIOR



Fonte: INSTAGRAM – Print de publicação no Instagram do Jogador Vinicius Junior.  
Disponível em: <https://www.instagram.com/vinijr/>. Acesso em: 10 de maio de 2024.

## ANEXO VII – PRINT DE PUBLICAÇÃO NO INSTAGRAM DO JOGADOR VINICIUS JUNIOR



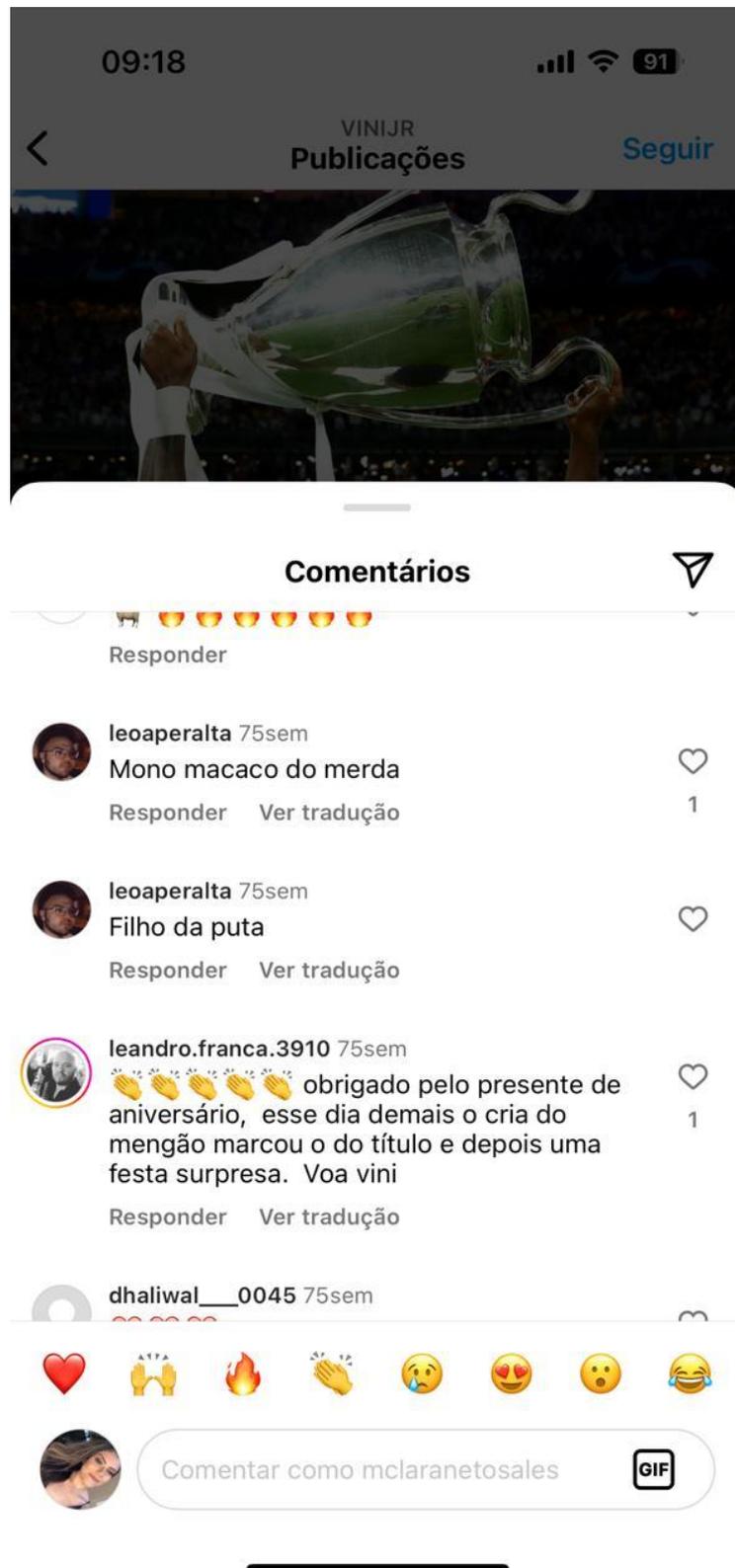
Fonte: INSTAGRAM – Print de publicação no Instagram do Jogador Vinicius Junior.  
Disponível em: <https://www.instagram.com/vinijr/>. Acesso em: 10 de maio de 2024.

## ANEXO VIII – PRINT DE PUBLICAÇÃO NO INSTAGRAM DO JOGADOR VINICIUS JUNIOR



Fonte: INSTAGRAM – Print de publicação no Instagram do Jogador Vinicius Junior.  
Disponível em: <https://www.instagram.com/vinijr/>. Acesso em: 10 de maio de 2024.

## ANEXO IX – PRINT DE PUBLICAÇÃO NO INSTAGRAM DO JOGADOR VINICIUS JUNIOR



Fonte: INSTAGRAM – Print de publicação no Instagram do Jogador Vinicius Junior.  
Disponível em: <https://www.instagram.com/vinijr/>. Acesso em: 10 de maio de 2024.