

CODIGO  
[nã**o** binário]

apoio



anatomia de uma

# ONDA DE ÓDIO

*Como a cumplicidade das plataformas transfere o ônus da violência a pessoas comunicadoras LGBTQIA+  
— o caso Entre Amíguas*

2026

## Prefácio

“Anatomia de uma Onda de Ódio” é um relatório sobre direitos humanos e integridade informacional no Brasil em 2024. Ele examina, com rigor metodológico e lente interseccional, a dinâmica de ataques coordenados, assédio e desinformação de gênero dirigidos ao Entre Amigos, podcast da Código Não Binário, e demonstra, a partir deste caso, os efeitos generalizados da ausência de regulamentação de plataformas, especialmente sobre comunicadoras(es) em direitos humanos e das pessoas LGBTQIA+ no Brasil. O documento parte do princípio de que ambientes digitais são espaços de vida pública, e quando a moderação falha e a violência fica livre — e é até impulsionada — ela transborda para o mundo físico, produzindo danos mensuráveis à saúde, à dignidade, e ameaçando até mesmo a integridade física dos grupos vulnerabilizados.

Nosso objetivo é duplo. Primeiro, **documentar**: descrever a cronologia da onda de hostilização, caracterizar as táticas e atores envolvidos (incluindo perfis que amplificam conteúdos sem necessariamente proferir xingamentos), e consolidar evidências. Segundo, **transformar evidência em ação**: avaliar a (in)efetividade de políticas de YouTube, Instagram, TikTok e X diante de violações claras, e apresentar recomendações operacionais para plataformas, Ministério Público e órgãos de governo, Legislativo e sociedade civil — com foco na proteção de pessoas trans e LGBTI+ e na redução do ônus imposto às vítimas.

O relatório organiza-se em cinco eixos:

**Linha do tempo analítica da onda de ódio**

**Fase 1: metodologia, cadeia de custódia e certificação de provas**

**Desempenho das Plataformas e Jornada de Denúncia**

**Fase 2: Análise de comentários, com base em métodos qualitativos e apoio de IA, incluindo leitura comparativa interplataformas**

**Conclusões e recomendações jurídicas e de política pública**

Assumimos três compromissos metodológicos: **responsabilidade**, que deriva do reconhecimento da posição situada e corporificada das pesquisadoras e das consequências de suas escolhas metodológicas, coleta de dados, análises e interpretações; **proteção**, resguardando a identidade de pessoas não públicas e minimizando riscos de reexposição; e **utilidade pública**, conectando achados a caminhos práticos de responsabilização e prevenção de danos, com o objetivo de promover a transformação social.

## Como citar e licença

---

Este relatório pode ser citado como:

Código Não Binário. *Anatomia de uma Onda de Ódio: Como a cumplicidade das plataformas transfere o ônus da violência a pessoas comunicadoras LGBTQIA+ — o caso Entre Amigas*. São Paulo, 2025.

O conteúdo está licenciado sob **Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International**.

Essa licença permite copiar, redistribuir e adaptar o material em qualquer meio ou formato, inclusive para fins comerciais, desde que seja dado crédito adequado à autoria original e indicadas eventuais modificações.

O objetivo é garantir **acesso aberto e circulação livre do conhecimento**, estimulando seu uso em pesquisas, políticas públicas, processos judiciais e ações de defesa dos direitos humanos.

## Agradecimentos

---

Este relatório é a versão expandida de um ensaio escrito em co-autoria da Código Não Binário — Veronyka Gimenes e Amanda Claro – com Giovanna Pignanelli, Horrara Moreira e Ocassana Côrtes. Estas pesquisadoras participaram das etapas de coleta e análise de dados da Fase 1.

O trabalho de coleta e análise de dados da Fase 1 foi possibilitado por parte de recursos emergenciais doados por Fundo Brasil de Direitos Humanos, Access Now, Fondo Acción Urgente para América Latina, Derechos Digitales e Digital Defenders Partnership. A etapa de inteligência artificial, análise de comentários da Fase 2 e a elaboração deste relatório foi viabilizada por parte dos recursos doados pelo Fundo Social Elas, através do edital Elas+ Cidadania Digital.

## RESUMO EXECUTIVO

Em 2024, o podcast Entre Amigues se tornou alvo de uma **mobilização coordenada de ódio online** que atravessou plataformas e transbordou para o mundo físico. O estopim ocorreu pouco mais de 48 horas após a publicação de um corte do 4º episódio da primeira temporada com o entrevistado Jupi77er, rapper transmasculino da periferia de São Paulo. Após rápida **viralização** do conteúdo no **Tiktok**, o **deputado federal de extrema-direita** Nikolas Ferreira replicou o vídeo em suas redes sociais em formato de *react* silencioso, sinalizando reprovação e convocando indiretamente seu público a falar por ele nos comentários. Em seguida, multiplicaram-se reações de terceiros, agora com ataques explícitos, e o termo **boyceta**, citado no trecho, permaneceu por dois dias entre os **Trending Topics no Twitter (X) no Brasil**.

A onda teve escala considerável: foram **mais de 12 mil comentários nos vídeos do podcast nos perfis do Entre Amigues** — a maioria com agressões, zombaria ou incitação à violência. Para entender o alcance e documentar violações, mapeamos conteúdos derivados (*remix*, *reacts*, citações) e testamos ferramentas de *social listening*. O processo expôs a **opacidade dos mecanismos de busca** — que retornavam resultados diferentes conforme o perfil logado — e a insuficiência das plataformas digitais hospedeiras e até mesmo das ferramentas comerciais para fins de responsabilização, o que nos levou a concluir a etapa com verificação manual e



classificação jurídica. Identificamos conteúdos estáticos e em vídeo, atribuídos a autores e “disseminadores”, e iniciamos a certificação de prova digital por serviços especializados para preservar evidências para responsabilização individual.

Ao acionar os mecanismos internos de denúncia, emergiu um padrão: as plataformas respondem com mais celeridade e previsibilidade a alegações de violação de direito autoral do que a denúncias de discurso de ódio. No **YouTube, denúncias por ódio não tiveram qualquer efeito**, e os vídeos denunciados permanecem ativos até a data de finalização deste relatório, sem medida efetiva. Já **por direito autoral, a plataforma abriu processo formal** com quatro possíveis desfechos e prazos definidos, embora apenas 1 de 22 vídeos tenha sido removido de imediato. Em diversas respostas, a plataforma deslocou o ônus probatório para a vítima, ancorando a análise em critérios de *fair use* do direito norte-americano — desalinhados com o ordenamento jurídico brasileiro. **O resultado é uma jornada longa, opaca e com negativas frequentes**, que impõe esforços desproporcionais às pessoas comunicadoras afetadas.

Os efeitos ultrapassaram as telas. A viralização odiosa teve consequências imediatas e intensas na vida, na saúde mental e nos relacionamentos das pessoas envolvidas, demandando medidas de segurança e mobilização de redes de suporte. **O caso também espelha tendências globais de campanhas anti-trans** que operam por *reacts* e códigos insinuados, terceirizando a agressão aos públicos e promovendo viralizações odiosas de conteúdos LGBTQIA+ e trans, na busca de **regular estes corpos através da violência digital** e de **mobilizar massas através do ódio a grupos minorizados**.

Os dados revelam uma **cadeia interplataformas**: o TikTok atua como gatilho de viralização, o YouTube consolida e perpetua a hostilidade, e o Instagram — embora com menor proporção — reencena o ataque por meio de memes, remixagens e comentários de chacota. O resultado é um **ciclo de amplificação simbólica e política do ódio contra pessoas LGBTQIA+, distribuído entre arquiteturas de plataforma distintas**.

Este relatório demonstra que a combinação entre **desenho de produto**, regras e **enforcement** atual das plataformas — com priorização prática de propriedade intelectual sobre dignidade — cria **barreiras estruturais à proteção de pessoas LGBTQIA+**. Propomos, ao final, recomendações para transparência, devida diligência

e responsabilização, com ênfase em processos de apelação legíveis, métricas públicas de moderação e canais de cooperação com a sociedade civil — caminhos já apontados por organizações do ecossistema de direitos digitais e LGBTI+.

## Linha do tempo da Onda de Ódio

Em maio de 2024, a Código Não Binário lançou o **Entre Amigos**, podcast independente que promove as vozes de pessoas LGBTQIA+, publicado semanalmente com cortes para redes sociais.

O quarto episódio foi apresentado em 22 de maio de 2025 nas redes por um corte de poucos segundos em que **Jupi77er**, rapper transmasculino paulistano, se apresenta com palavras que expressam suas dissidências de gênero e sexualidade, entre elas a palavra **boyceta**. O breve trecho, marcado por simplicidade e um sorriso no rosto, transmite conforto com o espaço de conversa conduzido por **Veronyka** e **Amanda** e com o conteúdo enunciado. O que inicialmente se supunha ser um corte de baixo apelo, em comparação com outros planejados, rapidamente **viralizou no TikTok**, com crescimento acentuado de visualizações de um dia para o outro. Já na quinta-feira, um dia após a estreia do episódio, equipe e convidado perceberam os sinais de **viralização odiosa**: aumento rápido de comentários hostis e uma razão de **compartilhamentos muito superior a curtidas**, indicador de circulação hostil voltada a “passar adiante” mais do que a expressar aprovação.

Pouco mais de **48 horas** após a postagem, em 24 de maio de 2024, o **deputado de extrema-direita Nikolas Ferreira** replicou o corte no Instagram e no TikTok em formato de *react* silencioso. Em um com a legenda “Melhor não”, em que faz expressões faciais e diz quase nenhuma palavra, sinalizou reprovação e induziu o público a concluir que “haveria muito a dizer” — ainda que não publicamente. Trata-se de uma estratégia de comunicação indireta recorrente no ambiente online: a figura pública **insinua** a reprovabilidade de um conteúdo e **terceiriza** a agressão aos comentários, preservando uma aparência de neutralidade e reduzindo riscos jurídicos. A leitura dos comentários ao vídeo permite concluir pelo sucesso da tática. Em 4 dias o vídeo já contava com quase 1,5 milhão de curtidas e mais de 200 mil comentários, como estes:

*"Se não se reproduzir, tá ótimo!"*, 28/05/2024, 831 curtidas, 35 respostas.

*"Satanás chega senta pra aprender"*, 24 de maio de 2024, 45.607 curtidas, 724 respostas, autoria de influenciador candidato pelo PL.

*"Faltou uma frase aí! Eu sou debioide"*, 24 de maio de 2024, 17.690 curtidas, 161 respostas, autoria de conhecida atleta brasileira.

*"Vim só ver o pessoal comentando, não tenho dinheiro pra advogado também. 😞"*, 29 de maio de 2024, 2 curtidas.

A partir dessa publicação, multiplicaram-se peças de terceiros, inclusive de outras figuras públicas como , , e . Entretanto, os novos conteúdos de terceiros passaram a ser principalmente de pessoas comuns e influenciadores digitais, agora com **ataques explícitos**, que reaproveitaram as imagens do entrevistado e das apresentadoras para produzir conteúdos similares aos do deputado. Ferramentas das próprias plataformas, como o Remix do Instagram, permitiram centenas de replicações do conteúdo do deputado, todas com o mesmo intento vexatório. O áudio do *react* do deputado — atingiu métricas consideráveis — que foi derrubado através de denúncia no Instagram fundamentada em direitos autorais.

Nesse período, **o termo "boyceta"** mencionado no trecho permaneceu por **dois dias** entre os **Trending Topics** do Twitter (X) no Brasil — onde o Entre Amigues não possuía perfil e não postou nenhum conteúdo diretamente —, sinalizando a escala da exposição e o efeito de arrasto para além da plataforma que originou o corte.

O encadeamento "reação silenciosa → comentários hostis em massa → conteúdos derivados" consolidou um padrão que este relatório descreve como operado por **disseminadores**: perfis de grande alcance que amplificam conteúdos e **ativam audiências** a dizer o indizível, maximizando engajamento via comentários e réplicas.

As consequências foram imediatas. A viralização odiosa impactou a vida, os vínculos e a saúde mental e física das pessoas envolvidas, exigindo a adoção de medidas de segurança e a mobilização de redes de suporte e segurança nacionais e internacionais. Em paralelo, organizou-se uma frente jurídico-probatória: **duas equipes** de advocacia e assistência jurídica iniciaram a documentação estruturada do

caso para fins de responsabilização, com preservação de evidências e preparação de medidas cabíveis.

## Fase 1: resposta rápida e triagem

### Metodologia

---

A investigação começou com perguntas operacionais: como localizar sistematicamente conteúdos ofensivos, como tipificar juridicamente as violações e como registrar prova útil à responsabilização individual dos ofensores. Organizou-se um **fluxo de trabalho** com reuniões periódicas da equipe para definir escopo, critérios e papéis. Nesse primeiro momento, optou-se por levantar manualmente comentários e publicações nos perfis do Entre Amigues no TikTok, Instagram e YouTube, por serem as plataformas diretamente utilizadas para distribuição dos episódios e cortes.

O X/Twitter foi deliberadamente excluído do escopo desta fase. Embora tenham sido identificadas publicações ofensivas, a ausência de perfil do podcast na plataforma e o ambiente regulatório e de moderação insuficiente para discurso de ódio reduziram a prioridade imediata de coleta de dados. A decisão preservou foco e recursos nos ambientes em que a equipe detinha presença editorial e, portanto, maior capacidade de ação.

Para organizar a triagem, adotou-se uma classificação inicial que orientou a partir do teor, a avaliação de encaminhamentos cíveis e criminais. Este esquema serviu como rascunho de *codebook* e será expandido e calibrado na análise de conteúdo sistemática apresentada adiante neste relatório.

No rastreio de conteúdos derivados (*remix*, *react*, citações), utilizou-se como palavra-sinal o termo **"boyceta"**. A equipe observou, entretanto, comportamento inconsistente dos mecanismos de busca internos, com retornos distintos conforme o perfil logado — um indício de opacidade algorítmica que compromete a reprodutibilidade da coleta e demanda auditorias por perfis de controle. Foi testada a ferramenta de *social listening*, mas não atendeu aos requisitos de precisão e cobertura, sendo descartada. O percentual de menções negativas apontado pela

plataforma (76%) é consistente com os achados da análise de conteúdo apresentada adiante. Retomou-se, então, a busca manual, complementada por varreduras em perfis públicos de parlamentares, figuras de direita e extrema-direita e influenciadores com histórico de violência de gênero online.

A consolidação dos achados desta primeira fase foi feita em planilha de acompanhamento, registrando data, link, natureza da peça (autor direto ou disseminador), plataforma, tipificação jurídica preliminar, além de campos para controle de denúncias, prazos e desfechos. Ao final de quase dois meses de trabalho, foram **selecionados conteúdos (estáticos e em vídeo) passíveis de ação cível e criminal na esfera individual, incluindo publicações de figuras públicas**. Esta avaliação não se deteve apenas à existência ou não de crime ou de dano indenizável — que certamente entregaria uma quantidade infinitamente maior de conteúdos —, mas também de viabilidade jurídica e conveniência da judicialização para as partes envolvidas, já que este é frequentemente um caminho de revitimização e risco sistêmico para as vítimas (SHECAIRA, 2025).

Em paralelo, iniciou-se o procedimento de preservação das evidências por serviços especializados. A equipe elegeu a Plataforma por critérios de idoneidade, confiabilidade, praticidade e custo, fundamentada nos arts. 158-A a 158-F do Código de Processo Penal e no entendimento do STJ sobre a inadmissibilidade de evidências digitais produzidas sem metodologia adequada. Ainda que tais custos permaneçam como barreira material relevante ao acesso à justiça para organizações e coletivos, a governança de dados observou os princípios de integridade, verificabilidade e minimização de risco.

Por fim, registram-se as **limitações** desta fase: a coleta manual está sujeita a deleções e edições retroativas; a **opacidade** dos algoritmos de busca compromete a replicabilidade; e ferramentas de *social listening* comerciais, embora úteis para exploração, não foram construídas com o propósito de mapear discurso de ódio não entregaram precisão e cobertura suficientes para fins de responsabilização neste caso. Essas limitações foram mitigadas por documentação tempestiva e pela triagem jurídica contínua dos achados, que orientou as denúncias e as decisões de preservação.



## Cadeia de custódia e prova digital

---

A responsabilização por discursos de ódio nas plataformas depende da forma como a prova digital é obtida, certificada e preservada. Em meios voláteis, sujeitos a edição e exclusão, a cadeia de custódia assegura integridade, autenticidade e rastreabilidade das evidências, disciplinada nos arts. 158-A a 158-F do Código de Processo Penal. A jurisprudência do STJ consolidou que provas digitais produzidas sem metodologia adequada são inadmissíveis no processo penal.

No caso do Entre Amigos, procedeu-se à certificação técnica por plataforma especializada, registrando contexto e temporalidade para demonstrar a dinâmica da escalada e a autoria das condutas.

Dado o **volume** típico de ondas de ódio, a preservação exigiu decisões de priorização para garantir representatividade e suficiência probatória: peças que inauguram a escalada; conteúdos de alto alcance; materiais que explicitem a tática de **disseminadores**; e amostras de comentários que evidenciem padrões de ofensa e convocação. Esse acervo foi indexado à planilha de acompanhamento descrita na Metodologia, com campos de data, link, autor, plataforma, tipificação preliminar e controle de denúncias e desfechos, preservando a rastreabilidade entre a evidência certificada e os passos de *enforcement* na plataforma.

Por fim, registramos que **custos e escolhas tecnológicas** — detalhados na seção metodológica — permanecem barreiras materiais ao acesso à justiça para coletivos e organizações pequenas. Ainda assim, a adoção de procedimento padronizado de coleta e certificação, aderente ao Código de Processo Penal e à orientação vigente do STJ, é imprescindível para sustentar medidas cíveis e penais e para dialogar com autoridades e plataformas em bases verificáveis.

## Desempenho das Plataformas e a Jornada de Denúncia

---

Após identificar autores e disseminadores e consolidar as evidências, foram acionados os mecanismos internos de denúncia de **YouTube, TikTok e Instagram**, em duas frentes: **discurso de ódio** e **direitos autorais**, permitindo comparar a eficiência das plataformas em dois campos de regulação distintos. Para monitorar o

percurso, a equipe manteve um registro sistemático de prazos, respostas e justificativas oferecidas por cada serviço. Foram denunciados nestas plataformas 98 conteúdos, sendo 54 no Youtube, 14 no Instagram e 30 no Tiktok.

O que emergiu, de início, foi um padrão consistente: as plataformas dedicam **mais celeridade e previsibilidade** às queixas por violação de copyright do que às denúncias por violência e transfobia. Essa assimetria, observada nos três serviços, informa o desenho desta seção, organizada pela experiência em cada plataforma.

A discrepância é mensurável: considerando o conjunto das 98 denúncias, as plataformas removeram **58% dos conteúdos por direito autoral**, mas apenas **8% dos denunciados por discurso de ódio**. Em todas as plataformas, a via autoral foi ao menos **50 pontos percentuais mais eficaz**. A diferença revela uma priorização sistemática da proteção de interesses econômicos sobre a integridade das pessoas usuárias.

No **YouTube**, a denúncia por discurso de ódio seguiu o fluxo padrão da plataforma: seleção do conteúdo infrator, escolha da categoria e indicação do trecho ofensivo. Em seguida, o sistema emitiu um e-mail de confirmação com **ID de denúncia** e os possíveis desfechos ("remoção", "restrições" ou "manter no ar").

No total, foram **54 denúncias** registradas, das quais **81,5% (44)** referiam-se a direito autoral e **18,5% (10)** a discurso de ódio. **Nenhuma das denúncias de ódio resultou em remoção efetiva (0%)**, e todas permaneceram com status "Recebida", sem análise final. Em contraste, **54,5% das denúncias autorais foram deferidas**.

**De acordo com a resposta do Youtube, os vídeos cujos trechos destacamos a seguir, não contém discurso de ódio:**

*"É óbvio que o afegão médio vai olhar pra isso e vai pensar: [entra outro vídeo, com homem fazendo uma careta e dizendo "você tá doido?"]. Não tem como levar a sério, não tem como algo assim ter minimamente qualquer credibilidade. O que que você tem a dizer sobre a boyceta, o boyceta, sei lá. Cara, assim, se você for classificar cada gênero como uma droga, ela chegou ao crack" – , canal Canal Tragicômico, 126.055 visualizações, publicado em 15 de junho de 2024.*

*"Tem gente falando que é palhaçada, tem gente incentivando, lado A, lado B, cada um tá com um lado. Eu tô do lado que eu não tô nem aí, eu vou falar a verdade, quer ser boyceta, eu sei que eu sou da tropa do machuca [fazendo movimento simulando um soco]" — , canal PAPO DE NEGÃO, 49.052 visualizações, publicado em 10 de junho de 2024.*

**Até a finalização deste relatório, em janeiro de 2026, todas as peças seguiam com status "Ativo"**, sem medida efetiva comunicada ao denunciante. A discrepância contrasta com a trilha de direitos autorais: há quatro vias formais de queixa, e, no formulário de remoção escolhido para simular a experiência de um usuário comum, o YouTube reconheceu a abertura do processo em até três dias e estabeleceu quatro desfechos operacionais (remoção pelo YouTube; remoção pelo próprio usuário; contestação do denunciado; ou pedido de informações adicionais).

Dos vídeos removidos por propriedade intelectual, apenas 1 de 22 foi removido de imediato; nos demais, a plataforma deslocou o ônus para a denunciante, ancorando a análise em "fair use" — doutrina norte-americana desalinhada com o direito autoral brasileiro. O resultado prático é um percurso burocrático, pouco transparente e desproporcionalmente custoso para as vítimas, mesmo quando há indícios claros de dano.

No **Instagram**, a trilha por discurso de ódio produziu, de forma quase sistemática, **respostas automáticas** informando que os conteúdos "não violavam as diretrizes", acompanhadas da recomendação para que as pessoas afetadas **deixassem de seguir** os perfis ofensores. Foram registradas **14 denúncias**, sendo **9 (64,3%)** por direito autoral e **5 (35,7%)** por discurso de ódio.

As taxas de remoção reforçam o padrão: **100% das denúncias autorais** foram efetivas, contra apenas **20% das denúncias de ódio**. As negativas vieram sem

critérios verificáveis ou justificativas substantivas, e as tentativas de apelação receberam devolutivas igualmente genéricas, o que inviabilizou contestações consistentes. Já as notificações por direitos autorais exigiram mais documentação e tiveram desempenho um pouco melhor, mas **sem um canal estruturado de acompanhamento**, o monitoramento do caso permaneceu opaco. Embora o Instagram tenha apresentado a maior taxa global de remoção (71,4%), o contraste entre 100% de eficácia em copyright e 20% em ódio confirma que o bom desempenho não se estende à proteção contra violência.

No **TikTok**, os obstáculos se repetiram. O formulário interno para denunciar conteúdos transfóbicos mostrou-se **limitado e de acesso pouco intuitivo**, e a plataforma **não oferece acompanhamento** do status das denúncias, o que impede verificar andamento e decisões. Foram registradas **30 denúncias**, das quais **21 (70%)** se referiam a discurso de ódio e **9 (30%)** a direito autoral. As taxas de remoção foram de **66,7% para autoral** e apenas **14,3% para ódio** — uma diferença de **52,4 pontos percentuais**. As respostas foram majoritariamente automáticas e sem indicação de providências. Na trilha de direitos autorais, o processo revelou-se ainda mais burocrático e, novamente, **referenciado no "fair use"** estadunidense, destoando do ordenamento brasileiro e impondo sucessivas complementações sem clareza sobre os critérios aplicados. O efeito combinado é um ambiente **fragmentado e opaco**, em que a remoção de conteúdos depende de procedimentos internos pouco acessíveis e de baixíssima previsibilidade.

O TikTok apresentou o **maior índice de negação de denúncias (70%)**, incluindo **85,7% de rejeição** às queixas por transfobia, o que reforça sua ineficácia em moderação reativa.

Ainda no Tiktok, até que a equipe descobrisse a funcionalidade de "Configuração de reutilização de conteúdo" em meados de junho, para impedir o uso do áudio e do vídeo em duetos e composições por outras pessoas, 1089 conteúdos foram gerados com o áudio do corte mais visualizado, a maioria deles em formato de *react* para chacota.

O conjunto dos dados demonstra uma política institucional de priorização econômica: os mecanismos de copyright operam com previsibilidade, prazos e retorno auditável. Já as denúncias por ódio são marcadas por opacidade, ausência de critérios e respostas automatizadas. A discrepância entre os dois regimes não decorre de falhas

técnicas, mas de uma **escolha estrutural de governança** que protege a monetização sobre os direitos humanos. A falta de transparência, a exigência de informações que não dialogam com o marco jurídico nacional e a ausência de prazos e justificativas auditáveis transferem para as vítimas o ônus de navegar burocracias, reunir provas e sustentar recursos — muitas vezes sem retorno substantivo.

Soma-se a isso uma limitação estrutural: as ferramentas de denúncia foram desenhadas para objetos **atomizados** (uma postagem, um comentário), mostrando-se impermeáveis à denúncia de **fenômenos coordenados** como a onda de ódio contra o Entre Amigues. Não é apenas humanamente inviável reportar manualmente milhares de mensagens hostis; ao fragmentar o problema em unidades isoladas, a própria ferramenta cristaliza a narrativa de que se tratam de episódios excepcionais e localizados — quando, na verdade, são campanhas amplificadas por arquiteturas de plataforma e atores organizados.

A ineficácia das denúncias de ódio também expõe os limites da moderação reativa: mesmo quando acionadas por vítimas diretas, as plataformas não removem conteúdos transfóbicos nem previnem sua recirculação. A combinação de prevalência altíssima de discurso de ódio e baixíssimas taxas de remoção revela a incapacidade — ou falta de vontade — das plataformas de exercer qualquer forma de diligência ativa.

À luz do julgamento do STF que declarou a inconstitucionalidade parcial e progressiva do art. 19 do Marco Civil e fixou parâmetros de diligência ativa das plataformas em hipóteses de circulação massiva de conteúdos gravíssimos (incluindo discurso de ódio), este caso ilustra o acerto dessa reinterpretção e a **urgência de operacionalizar** medidas de prevenção, detecção e resposta a campanhas coordenadas, para além da denúncia atomizada e do mero cumprimento reativo a ordens judiciais.

## **Fase 2: Análise de comentários**

Na fase 1, o trabalho de análise da onda de ódio realizado pela equipe dedicada focou inicialmente em certificar conteúdos em vídeo odiosos produzidos a partir de imagens e trechos do podcast e que ridicularizaram ou ofendiam as pessoas participantes.



Esta escolha inicial excluiu da análise os comentários feitos em postagens dos perfis do Entre Amigues e das pessoas participantes.

No entanto, **o volume de comentários e seu nível de toxicidade** — que tiveram um **impacto direto sobre a saúde mental e ativismo** das pessoas envolvidas — **não poderiam ser ignorados**. Em uma Fase 2 do projeto nos propusemos a analisar, especificamente, os comentários às postagens de episódios e cortes do podcast Entre Amigues em seus perfis no Instagram, Youtube e Tiktok. Nosso objetivo era **compreender a maneira como a onda de ódio se manifestou através deles**, se há padrões de discurso, uso de emojis, menções a conhecidos disseminadores e através de quais ferramentas discursivas a hostilidade se manifesta.

Fizemos uma **raspagem de dados** das postagens de cortes e episódios do podcast Entre Amigues em nossos perfis para criar as bases de dados a serem analisadas. Não foram explorados, nesta etapa, comentários em vídeos produzidos por terceiros, tampouco comentários em postagens do entrevistado. As bases contém comentários nas postagens desde 22 de abril de 2024, data da postagem dos primeiros vídeos teaser da temporada do podcast nas redes sociais e 22 de agosto de 2025, data da extração.

## Anotação manual

---

Uma anotação manual foi realizada na base de comentários do Instagram, que continha 2098 comentários. A raspagem não foi capaz de extrair o conteúdo dos comentários de resposta (replies), apenas a quantidade de replies, que foi de 808. Portanto, os comentários "replies" não tiveram seu conteúdo analisado.

A análise qualitativa dos comentários coletados nas publicações do *Entre Amigues* no Instagram teve como primeira etapa a atribuição de uma **avaliação geral**, sintetizada na variável *avaliação*. Essa categoria de topo representa a posição discursiva do comentário em relação ao conteúdo publicado — isto é, se ele expressa apoio, neutralidade, crítica legítima ou hostilidade. Essa triagem antecede a codificação das categorias identitárias e táticas (como transfobia, desumanização ou sexualização), servindo como uma forma de organizar o conjunto antes das análises mais específicas.

A abordagem analítica adotada no processo de anotação está fundamentada em uma perspectiva **foucaultiana** (Foucault, 1999a, 1999b), **feminista** (Haraway, 1988) e **queer** (Butler, 2003; Preciado, 2018, 2023), compreendendo a linguagem como prática social e como campo de disputa de poder. Partimos do entendimento de que os discursos não apenas descrevem a realidade, mas **a produzem**, instituindo verdades e regulando corpos. O discurso de ódio, portanto, é compreendido como **tecnologia disciplinar** — um instrumento de controle que opera pela repetição de normas cisheterossexuais, produzindo exclusão e hierarquia.

Em diálogo com os estudos feministas e queer, a pesquisa reconhece que toda produção de conhecimento é **situada e corporificada** (Haraway, 1988). As leituras realizadas nas etapas de anotação não se orientaram por uma pretensa neutralidade, mas por uma ética da escuta e uma atenção crítica às formas pelas quais a violência simbólica se manifesta nas interações digitais. Assim, a análise parte da vivência e do repertório social das pessoas anotadoras, compreendendo que a interpretação de um comentário não pode ser dissociada do contexto histórico e político em que se insere. Essa escolha epistemológica assegura que o processo de classificação traduza as dinâmicas de poder que estruturam a comunicação online e não apenas a literalidade das palavras.

Além disso, é importante esclarecer que **as noções de “ódio”, “LGBTfobia” e “transfobia” adotadas neste relatório extrapolam as definições estritamente jurídicas desses termos**. Quando empregamos essas expressões, não o fazemos como sinônimos dos tipos penais, mas como **categorias analíticas voltadas à compreensão dos fenômenos sociais que representam**. Reconhecemos que a própria tipificação penal dessas violências é produto de uma longa elaboração política e conceitual conduzida por movimentos sociais organizados e pela produção acadêmica das ciências humanas no Brasil e no mundo.

Nesse sentido, o presente estudo se alinha a essa tradição, **adotando uma posição epistemológica que não se limita às definições jurisprudenciais, por entendê-las insuficientes para abarcar a complexidade das violências vividas**. Além disso, entendemos este relatório como mais uma **ferramenta de mobilização política** no sentido do **necessário alargamento das definições destes conceitos**, na medida em que, em geral, **a jurisprudência** tem considerado atos individuais e episódicos, **deixando de captar as dinâmicas coletivas e coordenadas que caracterizam as**

**ondas de ódio nas plataformas digitais** e que foram observadas de forma paradigmática no caso Entre Amigues.

Este relatório tem caráter exclusivamente analítico e visa contribuir para a compreensão de dinâmicas sociais presentes nas interações estudadas. **Não se destina**, sob qualquer forma, **a imputar a pessoas físicas ou jurídicas a prática de crimes ou ilícitos em geral**, cuja apreciação compete às esferas competentes do Poder Judiciário. **As expressões 'ódio', 'LGBTfobia' e 'transfobia', portanto, são utilizadas em sentido sociológico e comunicacional**, para descrever formas de violência simbólica e estrutural observadas nas interações analisadas, sem pretender qualquer conotação jurídica ou acusatória.

Para esta codificação foi necessário compreender o contexto do vídeo, o que estava sendo discutido e quais identidades e corpos estavam presentes nas imagens e sons. As anotações foram realizadas pelas pessoas que produziram o podcast, o que facilitou a compreensão do contexto e da mensagem dos vídeos.

## Classificação geral dos comentários

A classificação geral foi estruturada em quatro categorias mutuamente exclusivas: **hostilidade/ódio, crítico sem hostilidade, neutro e positivo**.

A categoria **hostilidade/ódio** compreende comentários que expressam ataques explícitos ou implícitos a pessoas ou grupos, seja por meio de insultos, desumanização, negação de identidades, sexualização degradante, ameaças ou incitação à violência. Incluem-se aqui também os casos de escárnio, ironia ou “apito de cachorro” — expressões aparentemente neutras, mas que dependem de um contexto de referência partilhado para se configurarem como ataques. Essa categoria cobre tanto manifestações diretas (“isso é doença”, “porta do inferno abriu”) quanto formas veladas (“transformers”, “meus ovos”, “boyceta”), nas quais o conteúdo discriminatório é mediado por ironia, humor ou código cultural. Em todos esses casos, **o comentário é entendido como produtor ou reprodutor de violência simbólica**, e aparece associado a uma ou mais categorias identitárias (como **transfobia** ou **gordofobia**) ou táticas (como **assédio/insulto, desumanização** ou **patologização**). Posteriormente, na análise de severidade, essa categoria é subdividida em hostilidade (severidade 1) e ódio (severidade 2-3), conforme detalhado adiante.

A categoria **crítico sem hostilidade** abrange manifestações de discordância ou ironia que não se traduzem em ataques pessoais ou desqualificação de identidades. São expressões de dissenso legítimo — muitas vezes políticas ou morais — que preservam o reconhecimento da dignidade das pessoas envolvidas. Exemplo disso são comentários como “acho exagero”, “não concordo com tudo” ou “isso não muda nada”, que expressam desacordo, mas sem agressividade ou intenção de humilhar. Essa categoria é importante para evitar que o modelo confunda oposição ideológica com hostilidade.

A categoria **neutro** reúne os comentários que não expressam uma posição avaliativa clara. Incluem-se aqui perguntas, observações factuais, repetições de expressões do vídeo, interjeições e reações automáticas (“kkk”, “hahaha”, “😂”) quando desprovidas de valência negativa. Também entram nessa categoria expressões curtas e ambíguas que, isoladamente, não permitem inferir intenção de apoio nem de ataque. A função dessa categoria é reduzir falsos positivos, garantindo que o ruído natural das interações em rede — especialmente no uso de risadas e emojis — não seja interpretado como hostilidade.

Por fim, a categoria **positivo** inclui manifestações de apoio, empatia, solidariedade ou celebração em relação ao conteúdo do *Entre Amigues*. São expressões de afeto, concordância e defesa, como “amei”, “é isso”, “perfeitas ❤️ 🏳️‍🌈”, “respeitem as pessoas trans”. Também abrange comentários de contradiscurso, nos quais o público responde a ataques com mensagens de acolhimento ou resistência. Na base de dados, comentários que pedem o envio de link para o podcast por automação em mensagem direta (DM) também foram classificados como positivos ou neutros, conforme o caso.

Essas quatro categorias compõem, portanto, o primeiro eixo da análise — uma leitura discursiva de cada comentário enquanto ato comunicativo. A distinção entre elas foi fundamental para garantir a precisão das análises subsequentes, permitindo identificar quando um comentário é **de hostilidade/ódio**, **apenas crítico**, **neutro** ou **de apoio**.

É importante esclarecer a distinção terminológica adotada neste relatório entre “hostilidade” e “ódio”. **Hostilidade** é o termo mais amplo, que designa qualquer forma de agressão, desrespeito, ataque ou comportamento negativo dirigido a pessoas ou grupos. Inclui desde formas leves de desrespeito até ameaças graves, mas não permite distinção de níveis de severidade sem análise qualitativa adicional. **Ódio**, por

sua vez, é uma subcategoria específica de hostilidade, definida como comentários que expressam ataques severos à dignidade, desumanização, negação de identidades, sexualização degradante, ameaças ou incitação à violência. A classificação como "ódio" requer análise qualitativa com níveis de severidade (2-3) para precisão analítica.

No Instagram, onde foi realizada análise manual completa com codificação temática e níveis de severidade, utilizamos o termo "hostilidade" para designar comentários de severidade 1 (ofensa leve/deboche) e "ódio" para severidade 2 (ataque severo à dignidade) e severidade 3 (ameaça ou incitação). No TikTok e YouTube, onde a análise foi realizada por modelo de IA com classificação binária (sem níveis de severidade), utilizamos o termo "hostil" ou "hostilidade" para refletir a limitação metodológica: embora o modelo identifique comportamentos hostis, não é possível, na versão atual, distinguir níveis de severidade dentro dessa categoria. Esta distinção terminológica preserva a precisão analítica e a transparência metodológica, reconhecendo que "hostilidade" é um termo válido e amplo que inclui "ódio", mas não permite a granularidade analítica alcançada na classificação manual do Instagram.

## Codificação temática e níveis de severidade

Após a triagem inicial dos comentários segundo sua **avaliação geral** procedeu-se à **anotação temática e tática**, isto é, à identificação do tipo específico de violência ou atitude discursiva presente em cada comentário. Essa segunda camada da anotação permitiu diferenciar entre tipos discursivos, distinguindo entre ataques dirigidos a identidades específicas e modos mais amplos de hostilidade.

As categorias de codificação foram divididas em dois grupos: **(1) categorias identitárias**, que indicam o grupo social ou identidade de gênero/sexualidade alvo do ataque; e **(2) categorias táticas**, que descrevem a forma discursiva assumida pelo ataque.

### Categorias identitárias

As categorias identitárias correspondem aos grupos sociais visados nas mensagens. Cada comentário podia receber um ou mais desses marcadores, de acordo com o conteúdo expresso. As categorias utilizadas foram:



Transfobia	ataques, piadas, deslegitimação ou desumanização de pessoas trans e travestis; inclui misgendering e deboche com linguagem neutra.
Homofobia, Lesbofobia, Bifobia	ofensas ou estigmatizações dirigidas a orientações sexuais específicas.
Intersexofobia	ataques ou ironias sobre corpos intersexo.
Gordofobia	ofensas, ridicularização ou metáforas pejorativas sobre corpos gordos.
LGBTfobia geral	categoria genérica utilizada apenas quando o ataque se dirige à comunidade LGBTQIA+ de forma difusa, sem especificar o grupo (por exemplo, "esse povo", "essas coisas").

Quando um comentário apresentava transfobia explícita, não se marcava simultaneamente a categoria genérica de LGBTfobia, para evitar duplicidades e garantir maior precisão analítica. Inicialmente foram consideradas as categorias racismo, intersexofobia e intolerância religiosa, pois supomos que poderiam ocorrer em razão dos temas tratados nos vídeos, mas não houve quantidade de comentários suficientes nestas chaves analíticas para garantir dados robustos, levando-nos a desconsiderá-las.

## Categorias táticas

As categorias táticas descrevem o modo discursivo do ataque — ou seja, a maneira pela qual a hostilidade e o ódio se manifestam linguisticamente. Foram as seguintes:

Misgendering/Deadnaming	Negação do gênero autodeclarado, uso de pronomes e concordância de gênero incorretos, nome de registro ou expressões que reduzem a identidade de gênero à genitália.
Desumanização/Animalização	Redução da pessoa a um ser não humano, objeto ou animal ("esse ser", "bicho", "coisa", "dinossauro", "merda").
Ameaça/Incitação	Mensagens que desejam ou convocam violência física, extermínio ou expulsão ("deveria ser extinto", "não sobrará um").

Assédio/Insulto	Ofensas diretas, xingamentos, escárnio e zombaria, incluindo sarcasmo e deboche
Sexualização/Assédio sexual	Uso de expressões de teor sexual degradante, invasivo ou de incitação a violência sexual.
Pânico moral e crianças	Narrativas que associam identidades de gênero e sexualidade à corrupção moral ou à ameaça a menores
Patologização/Pseudociência	Uso de termos médicos, psiquiátricos ou biológicos para enquadrar identidades de gênero e sexualidade como doença, desvio ou anomalia
Desinformação de gênero	Afirmações falsas sobre identidades trans e dissidentes (por exemplo, "só existem dois gêneros", "isso é moda").

Essas categorias podiam coexistir em um mesmo comentário, permitindo capturar a complexidade e a interseccionalidade das violências discursivas.

### Escala de severidade (0–3)

Cada categoria (identitária ou tática) foi avaliada segundo uma **escala ordinal de severidade**, que mede o grau de violência simbólica expressa:

0 – Ausente ou neutra	Não há hostilidade; o conteúdo é informativo, positivo ou contradiscursivo
1 – Hostilidade (ofensa leve/deboche)	Ridicularização, ironia, risadas ou linguagem pejorativa leve, não explicitamente violento, e onde não era possível identificar um destinatário específico
2 – Ódio (ataque severo à dignidade)	Negação de identidade, desumanização, patologização, sexualização degradante ou difamação explícita onde havia um destinatário do ódio — pessoa ou grupo específicos
3 – Ódio (ameaça ou incitação)	Desejo de morte, eliminação, expulsão, violência física ou sexual.

Por fim, foi calculada uma severidade global para cada comentário, a partir do valor máximo entre todas as categorias codificadas (identitárias e táticas).

Assim, se um comentário apresentasse "transfobia = 2" e "desumanização = 3", sua severidade global seria 3. Esse indicador resume o potencial de dano discursivo de cada comentário, permitindo análises quantitativas e comparações entre plataformas, publicações e momentos da onda de ódio.

Essa metodologia de codificação manual, fundamentada em leitura contextual e interpretação situada, garantiu que a classificação refletisse não apenas a literalidade das palavras, mas também os sentidos sociais e culturais implicados nos comentários. O resultado é um corpus anotado capaz de sustentar análises tanto qualitativas quanto computacionais sobre o funcionamento contemporâneo do discurso de ódio digital contra pessoas LGBTQIA+.

## Treinamento de IA

---

No Youtube e no Tiktok a quantidade de comentários era muito superior. A raspagem retornou 9998 comentários em ambas somadas, tornando inviável a anotação humana considerando os reduzidos recursos humanos, financeiros e psicológicos da organização e da equipe, além das construções temporais — jurídicas e políticas — que exigiam um resultado mais rápido. Diante destas limitações, e projetando a utilidade da replicação desta metodologia em outros contextos e por outras ativistas, decidimos pela criação de um modelo de IA.

A base de comentários do Instagram anotada manualmente serviu como corpus de **treinamento para um modelo de linguagem** voltado à detecção de discurso de ódio em português, o . O modelo foi desenvolvido a partir do Tupi-BERT base portuguese-cased, um BERT pré-treinado em português, posteriormente ajustado (fine-tuned) para identificar especificamente o ódio contra pessoas LGBTQIA+. O modelo, nesta versão, faz uma identificação binária hostil/não hostil. É importante notar que, embora o modelo tenha sido treinado com exemplos de "ódio" (categoria específica com severidade), sua classificação binária não permite distinguir níveis de severidade dentro da categoria "hostil". Portanto, para TikTok e YouTube, utilizamos o termo "hostil" ou "hostilidade" — mais amplo e metodologicamente adequado — em vez de "ódio", que requer análise qualitativa com níveis de severidade.

O treinamento utilizou Focal Loss ( $\alpha=0.75$ ,  $\gamma=2.0$ ) e threshold 0.30. Os resultados demonstram alta capacidade de detecção: recall de 98,44%, acurácia de 76,15%,

precisão de 61,13% e F1-score de 75,42%. Embora cerca de 23% dos alertas sejam falsos positivos, o modelo prioriza deliberadamente a sensibilidade para evitar que comportamentos hostis passem despercebidos.

Como **limitações e salvaguardas**, registra-se que o modelo foi treinado em **português brasileiro** e em **domínio Instagram**, podendo ter desempenho inferior em outras plataformas e variações linguísticas do português. Ironia e sarcasmo seguem desafiadores e explicam parte dos **falsos positivos (~23%)**, motivo pelo qual se criou uma fila de revisão humana para os conteúdos marcados como nível de confiança médio, abaixo de 0,6.

A classificação dos comentários de TikTok e YouTube foi realizada pelo modelo de IA, com revisão humana exclusivamente para casos de confiança inferior a 0,6. Do total de 10.004 comentários, 3.157 (31,6%) passaram por revisão humana, garantindo maior precisão nos casos limítrofes.

A confiança média do modelo foi de 0,685, sendo 0,649 para comentários classificados como HOSTIL e 0,786 para NÃO-HOSTIL, indicando que o modelo é conservador ao classificar hostilidade. Dos 6.847 comentários com confiança  $\geq 0,6$  (68,4%), que não foram revisados, a classificação automática foi mantida. Notavelmente, entre os 3.157 comentários revisados humanamente (confiança  $< 0,6$ ), 92,1% foram confirmados como HOSTIL, validando a tendência conservadora do modelo. Os resultados de 70,5% (TikTok) e 80,4% (YouTube) de comentários hostis representam valores conservadores que refletem um cenário mínimo, não máximo, de hostilidade nas plataformas.

## Análise de conteúdo

---

### Instagram

A base de comentários raspada dos vídeos do Entre Amigues no Instagram, anotada manualmente, foi submetida a análise e processamento através de scripts em linguagem de programação Python que extraíram estatísticas descritivas, análises temporais diárias, distribuições de severidade, padrões textuais agrupados e análise de emojis. Os pontos de atenção para garantir confiabilidade incluíram: validação de

consistência entre diferentes métodos de contagem, verificação de parsing de campos com quebras de linha, confirmação manual de números críticos exemplificativos, e verificação cruzada entre dados brutos e análises agregadas.

A análise documenta um **ambiente de interação marcado por hostilidade sistemática contra pessoas LGBTQIA+** e, em particular, **contra pessoas trans**. Dos dados é possível observar como a **conversação pública** se organiza para **produzir diferença e reduzir legitimidades** por meio de rotinas expressivas que vão **do insulto à incitação à violência**, transitando por zonas cinzentas de **deslegitimação cotidiana**.

No plano quantitativo, um terço da conversa se ordena em torno de práticas de ataque: 33,56% do total são comentários hostis ou odiosos (704), sendo estes classificados como hostilidade (severidade 1) ou ódio (severidade 2-3), ante 961 positivos, 190 neutros e 198 críticos sem hostilidade. Comentários marcados como "Not Available" — apenas 2,1% — sugerem ter havido algum tipo de remoção, seja pelo usuário ou pela plataforma, ou mesmo invisibilidade do comentário por bloqueio.

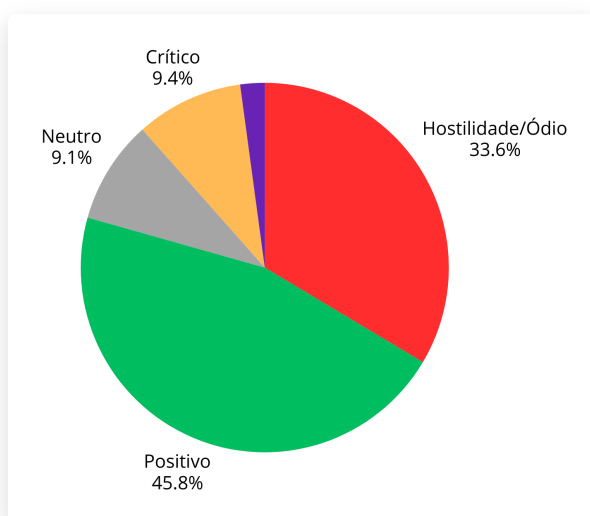


Gráfico 1

A severidade média global é 1,378 (escala 1–3), com concentração em severidade 1 (hostilidade) e severidade 2 (ódio) e um núcleo menor, porém insistente, de severidade 3 (ódio), o que indica um **contínuo de coerção discursiva que naturaliza a hostilidade e abre espaço para picos de intensidade explícita de ódio**.



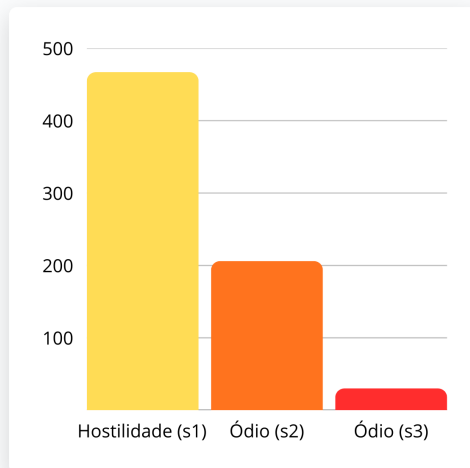


Gráfico 2

A série temporal adiciona contexto a esse volume. O relatório marca duas datas — 24/05/2024 e 27/06/2024 — como linhas de referência que atravessam os gráficos de evolução por tipo, percentual de hostilidade/ódio e severidade média. A primeira corresponde à data em que o deputado de extrema-direita Nikolas Ferreira realiza postagem no instagram reagindo ao vídeo de corte do podcast em que o entrevistado Jupi77er se anuncia como "boyceta". A segunda marca a data da postagem do último corte do podcast na temporada.

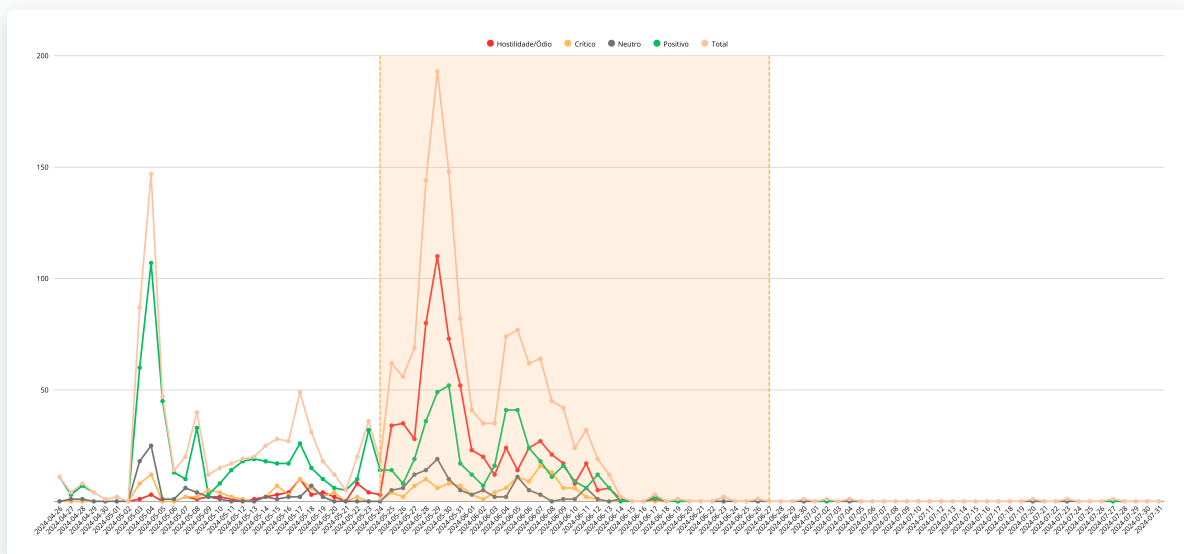


Gráfico 3

Observe-se que, **antes da postagem do deputado em 24/05, o perfil do podcast já recebia um grande número de comentários**, atraídos provavelmente pela imagem e influência de Guilherme Terreri, educador e criador da drag queen Rita Von Hunty, convidado do primeiro episódio. Apesar de ser muito conhecido e ter trazido temas de gênero e sexualidade considerados polêmicos nas redes sociais, como linguagem neutra e demissexualidade, os comentários hostis ou odiosos (severidade 1-3) eram, em média, 7,24% do total global. **Após o deputado visibilizar o conteúdo à sua base, amplificando a viralização, a curva sobe a 47,21%**, representando um **aumento de 552% na proporção de hostilidade e ódio**.

Isso significa que a relação entre o volume de comentários hostis/odiosos e o volume geral de comentários aumentou drasticamente, passando de aproximadamente 1 em cada 14 comentários para quase 1 em cada 2 comentários sendo hostis ou odiosos. A visualização a seguir nos mostra que a partir de 24/05/2024, os comentários hostis e odiosos passam a superar os comentários positivos, o que não havia acontecido antes.

A proximidade entre esses marcos e elevações nas curvas reforça a leitura de que **a visibilidade pública e a chamada à ação construída pelo vídeo do deputado constitui um disparador de fala ofensiva**: quando certos conteúdos se tornam mais dizíveis, a gramática do ataque encontra menos atrito para se repetir. A análise não afirma causalidade mecânica, mas demonstra **contiguidade temporal suficiente para sustentar que as janelas de maior atenção coincidem com recrudescimento do conflito**.

No recorte de alvo, **a transfobia aparece como eixo de organização do dano**. São 203 comentários de cunho transfóbico, distribuídos entre 115 em severidade 1, 87 em severidade 2 e 1 em severidade 3, combinação que revela persistência de práticas de negação identitária, desqualificação de experiência vivida e policiamento das fronteiras do “reconhecível”, em vez de episódios isolados sem padrão. É nesse

registro que **as pessoas destinatárias são interpeladas reiteradamente a caber ou não caber na moldura do humano e do legítimo.**

Os dados relativos às subcategorias táticas — assédio/insulto, patologização/pseudociência, misgendering/deadnaming, desinformação de gênero, ameaça/incitação, sexualização/assédio sexual e pânico moral/crianças — descrevem um mesmo mecanismo de produção de dano operando por registros distintos, mas complementares. No agregado, elas ajudam a entender por que a hostilidade não é apenas volumosa: ela é organizada, tem repertórios estáveis e usa linguagens que emprestam aparência de autoridade a julgamentos morais sobre corpos e identidades.

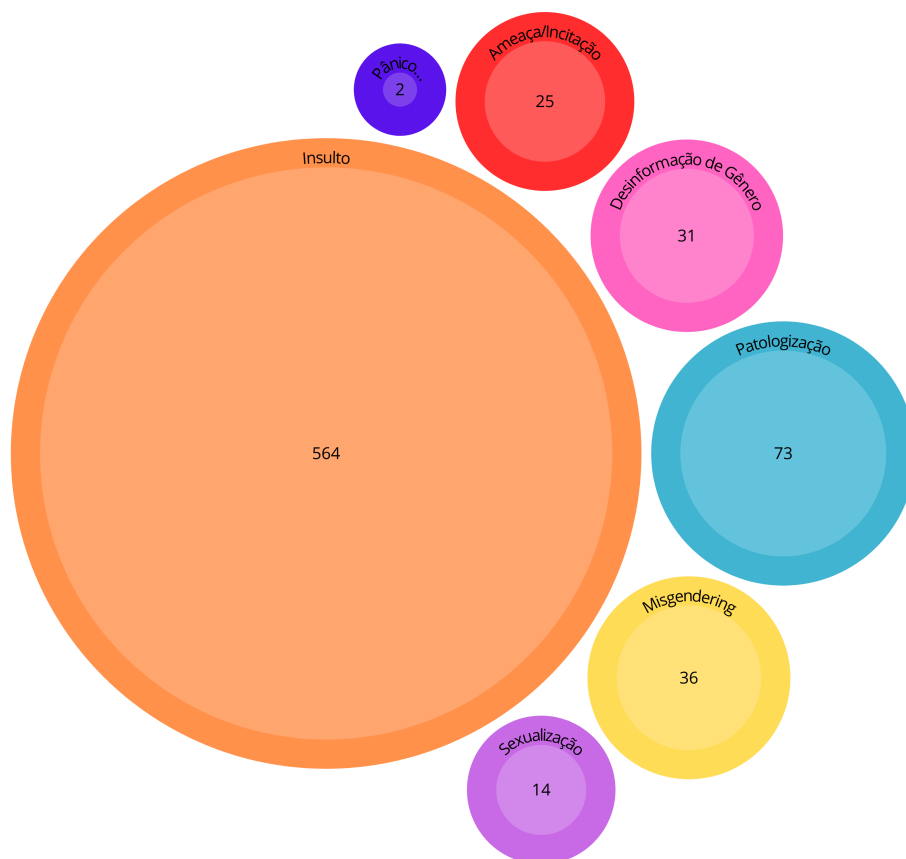


Gráfico 4

A categoria tática assédio/insulto emerge como a forma mais prevalente, representando 564 comentários (80,1% dos comentários hostis/odiosos; 26,88% do total), com severidade média de 1,09. A patologização/pseudociência é o segundo eixo mais prevalente (10,38% dos comentários hostis/odiosos, severidade média 1,59), funcionando como dispositivo de "tradução técnica" da deslegitimação. O misgendering/deadnaming aparece com 36 comentários (1,72% do total), mas concentra a maior intensidade relativa (severidade média 1,92), com expansão de 3 para 33 casos após 24/05/2024 (+1.000%). A desinformação de gênero combina baixa frequência (31 comentários; 1,48% do total) com altíssima intensidade (severidade média 1,94), com aceleração temporal expressiva de 2 para 29 casos após 24/05/2024 (+1.350%).

As categorias táticas apresentam ainda um comportamento de encadeamento com outras formas de violência, especialmente com transfobia. **As co-ocorrências entre categorias táticas e identitárias indicam que a violência não aparece atomizada, mas encadeada em "pacotes" que combinam humilhação, pedagogias de desinformação e vernizes de cientificidade.** Entre as mais frequentes, destacam-se os acoplamentos de assédio/insulto com zombaria via emoji, de assédio/insulto com transfobia e de assédio/insulto com patologização/pseudociência, além de interseções entre desinformação de gênero, misgendering/deadnaming e transfobia. É essa composição que produz densidade: uma piada, um riso gráfico, uma suposta explicação médica e uma reclassificação jurídica se somam para transformar diferença em inferiorização plausível aos olhos de uma audiência. A seguir, exemplos de comentários por categorias identitárias e táticas prevalentes.

Comentário	Categorias e severidade
"@entre_amigues @jupi77er vocês são engraçados é tipo o Mussun o cebolinha que falam errado. Não binário queer boyceta amigues todes TODA ESSA PALHAÇADA NÃO EXISTE máximo respeito a pessoa de vocês só façam um favor pra sociedade não tenham filhos só isso já tá bom pois assim daqui 20 anos essa palhaçada não terá sequência e o Brasil volta ao normal"	Insulto (s1), desinformação de gênero (s2)

"Vc seria A G0rd0NA da Hornet?"	Insulto (s2), gordofobia (s2), misgendering (s2)
""O hospício que vc fugiu já está chegando na sua porta aí com camisa de força 😂😂😂😂😂 #nikolasferreiraomelhordobrasil"	Patologização (s3)
"Tem que tomar uns tapão na orelha pra para de maluquice"	Incitação a violência (s3), patologização (s1)
""Os próximos gêneros: Transtornade, esquizofrênique, epilétique entre outros"	Insulto (s1), patologização (s2), transfobia (s1)
"É só uma lésbica com esquizofrenia."	Patologização (s1), lesbofobia (s1), misgendering (s2)
"Na comunidade de vcs não tem transDOENTE não? 🤔🤔🤔 "	Transfobia (s1), patologização (s2)
"E por esse mesmo motivo, pessoas acéfalas falam AMIGUES, TODES, ILE, DILE... e nós precisamos ouvir e aceitar ainda, o que é pior."	Transfobia (s1), patologização (s1)
"Gente, minha mãe é psicóloga do caps, disse que a boyceta fugiu e esta tendo alucinações, fiquem longe, essa doença é transmissível. A polícia já está procurando ela 🙌"	Transfobia (s2), misgendering (s2), patologização (s3)
"Boyceta gay essa coisas aí, são a decepção do mundo, 🧐"	Transfobia (s2), homofobia (s2),
"Isso é falta de ro la no meio do ra bo !"	Sexualização (s3)

A camada semiótica mostra por que essa gramática circula com tanta eficiência. Quase um terço dos comentários contém emoji (29,69%; 623 mensagens), totalizando 1.475 ocorrências e 156 tipos únicos; nas mensagens de ataque, sequências do emoji de riso e signos como 🤔 funcionam como atalhos de zombaria e



desumanização, codificando concordância e reforçando laços entre pares com baixo custo de produção de agressão. A seguir alguns exemplos ilustrativos do uso de emojis por categoria anotada.

Comentário	Categoria de emoji
"Falta de internação em algum manicômio 🤡 🤡 "	Desumanização
"🤡 🤡 🤡 "	Desumanização
"Eu sou boypica amigues 😂😂😂😂😂😂😂😂😂"	Zombaria
"Daqui a pouco vão criar a sigla LGTV HDMI pessoal que se identifica como eletrodomésticos 😂😂😂"	Zombaria
"TODES É MEU OVOS !!! 😂😂😂😂😂😂😂😂😂"	Zombaria
""Eu queria 1 ano do Che e do Fidel no Brasil .... Não sobrava um 🔥 "	Violência

A economia do engajamento explica por que esse conteúdo se torna ubíquo: **comentários hostis e odiosos têm maior probabilidade de receber replies (15,2%) do que os positivos (12,8%)**, transformando ofensa em circulação e visibilidade. Esse mecanismo faz com que a **violência** seja não apenas possível, mas **produtiva em termos de alcance**.

Tomado em conjunto, o material descreve um **ambiente em que a combinação de volume, encadeamentos táticos e janelas de alta visibilidade produz risco previsível e evitável às pessoas LGBTQIA+, sobretudo às pessoas trans**. Em outras palavras, os dados não apenas contam uma história de hostilidade sistemática; também apontam, com precisão suficiente para tutela coletiva, onde estão os pontos de intervenção para reduzir dano, redistribuir o ônus do cuidado e restabelecer condições mínimas de fala.

## Youtube

No Youtube, a análise foi binária — hostil ou não hostil — sem subcategorias táticas e identitárias, na medida em que foi inicialmente realizada pelo modelo de IA descrito

anteriormente. Nos vídeos do Entre Amigos nesta plataforma, foram raspados 3.733 comentários entre 16/04/2024 e 11/08/2025.

O YouTube se destaca não apenas pela alta prevalência de hostilidade (80,4%), mas pela **persistência e consistência desse padrão ao longo de 15 meses**, com comentários registrados em 442 dias distintos.

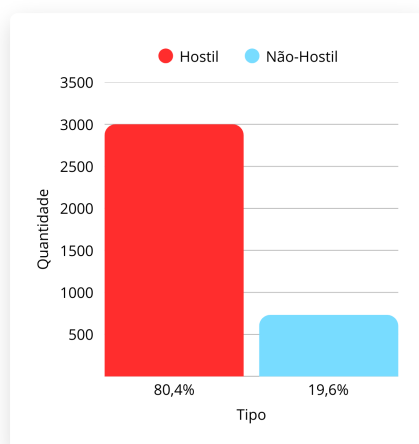


Gráfico 5

Diferentemente das demais plataformas, a hostilidade aqui não se comporta como uma reação episódica, mas como **um ambiente contínuo**. Antes da publicação do vídeo de Nikolas Ferreira, em 24 de maio de 2024, apenas 38,2% dos comentários expressavam hostilidade. Após a viralização do ataque, essa proporção saltou para 83,7%, mantendo-se acima de 80% nos meses seguintes.

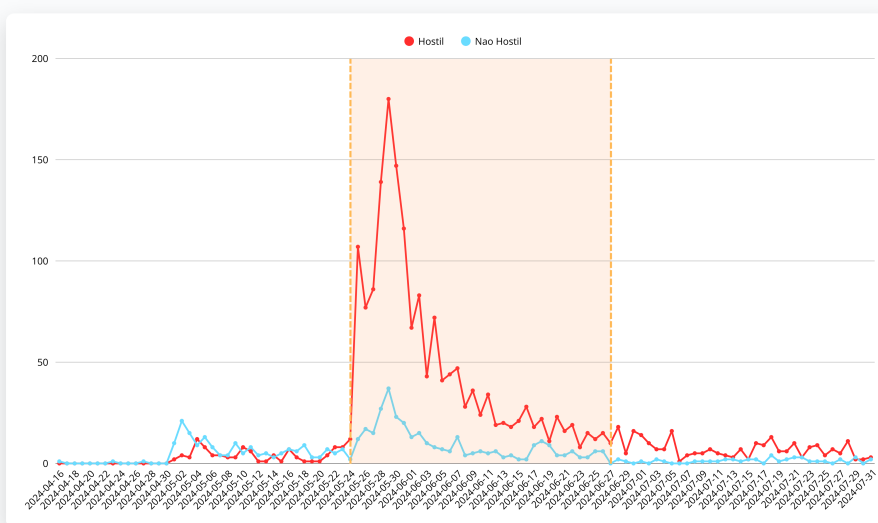


Gráfico 6

Esse crescimento progressivo indica **um processo de normalização da hostilidade**, sustentado pela inércia dos mecanismos de moderação e pela lógica de recomendação da própria plataforma. O YouTube não apenas falhou em conter a onda inicial, mas **preservou e alimentou a circulação dos comentários ofensivos** mesmo depois do encerramento das postagens do podcast, revelando a autonomia do ódio em relação à atividade do perfil afetado.

O caráter “autossustentado” dessa hostilidade sugere que o YouTube opera como um **arquivo ativo de violência**, no qual o engajamento negativo é recompensado por visibilidade. O resultado é um ecossistema em que mais de oito em cada dez comentários reproduzem insultos, estigmas de gênero e narrativas transfóbicas — um cenário de deterioração gradual do espaço discursivo, que combina **longevidade, intensidade e falha estrutural de moderação**.

## Tiktok

No Tiktok, a análise foi binária por modelo de IA nos mesmos moldes do Youtube. Nesta rede, foram raspados 6.265 comentários considerando o mesmo período, mas não há novos comentários após 18/12/2024.

Nesta rede, o padrão é distinto. A plataforma concentrou **a maior quantidade absoluta de comentários (6.265)**, mas o comportamento foi explosivo e efêmero. A onda começou **dois dias antes** da postagem de Nikolas Ferreira, em 22 de maio de 2024, com 3.025 comentários em apenas dois dias, dos quais 70,8% eram hostis. Essa **antecipação temporal** em relação às demais plataformas, sugere estar correta a constatação de que foi, de fato, no Tiktok que a onda de viralização hostil se iniciou. No Tiktok, o conteúdo já havia circulado de forma orgânica ou algorítmica, potencialmente impulsionado por mecanismos de recomendação automática, levando o conteúdo mesmo para pessoas não LGBTQIA+ e que repetidamente afirmavam, em tom de chacota, não entender o que estava sendo falado no vídeo.

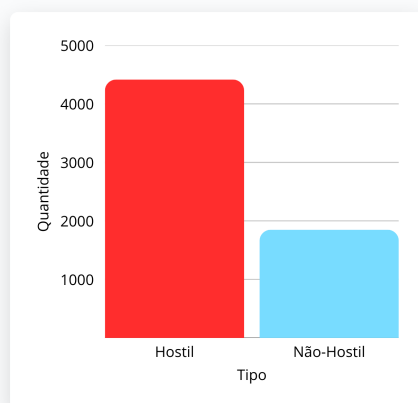


Gráfico 7

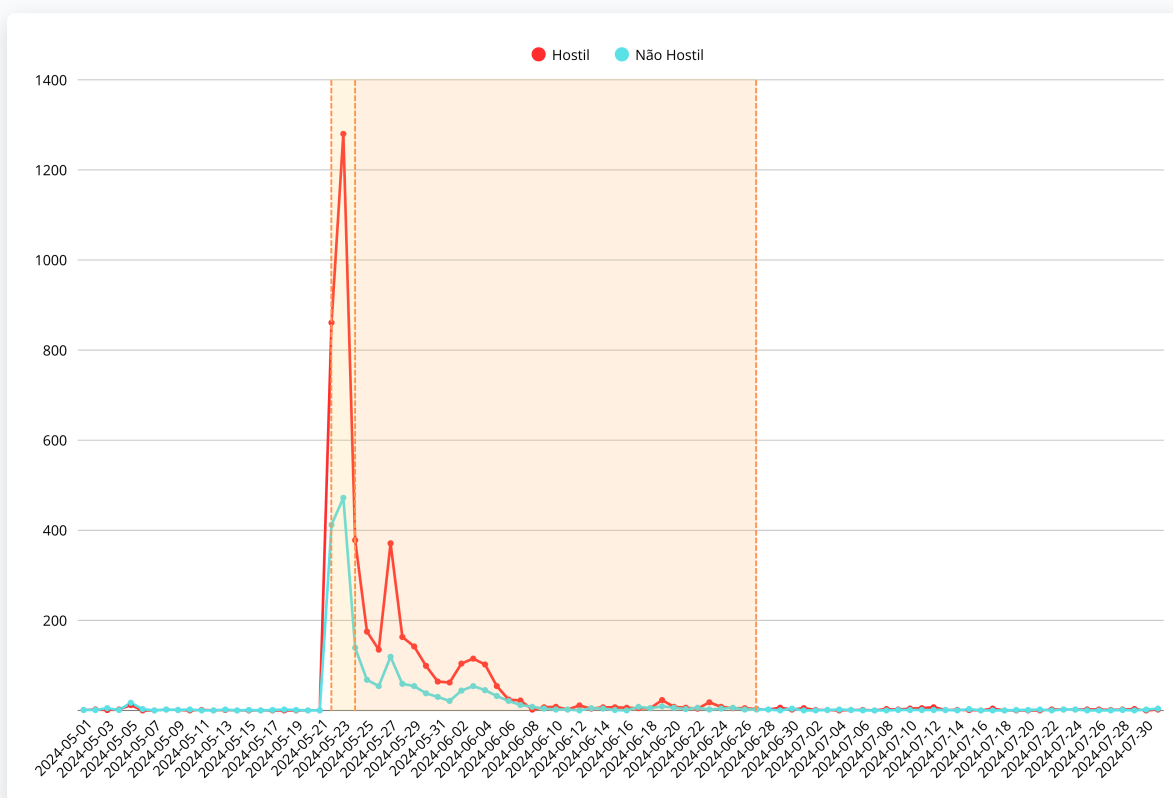


Gráfico 8

O TikTok evidencia, assim, **a dinâmica de viralização instantânea e esquecimento rápido** típica da plataforma. A hostilidade explode de maneira concentrada e intensa, mas não se sustenta no tempo. Essa forma de circulação, ainda que breve, tem alto impacto: em poucos dias, conteúdos transfóbicos atingiram milhões de visualizações, estabeleceram as condições para construção de um factóide transfóbico para

mobilização política de pânico moral por personalidades de extrema-direita, que por sua vez produziram seus conteúdos relacionados, levando o tema para as demais redes onde ele não teria viralizado por si só — e onde o podcast sequer possuía perfil, como no Twitter (X) –, consolidando uma narrativa estigmatizante multi-redes. O padrão combina **velocidade e amplitude**, revelando a vulnerabilidade das vítimas frente a uma arquitetura de difusão que privilegia a reatividade e o choque emocional.

## Análise comparativa entre plataformas

A comparação entre YouTube, TikTok e Instagram revela **três arquiteturas distintas de amplificação da hostilidade**, que, articuladas entre si, **produzem uma ecologia de violência digital contínua e interdependente**. Ainda que cada serviço opere com públicos e algoritmos diferentes, o conjunto dos dados mostra um fenômeno coerente: a **ascensão coordenada do discurso transfóbico** em reação à visibilidade do podcast *Entre Amigues*.

Em termos quantitativos, a prevalência de hostilidade variou de 34,3% no Instagram a 70,5% no TikTok e 80,4% no YouTube, também havendo variações significativas de volume, como sugere o gráfico a seguir.

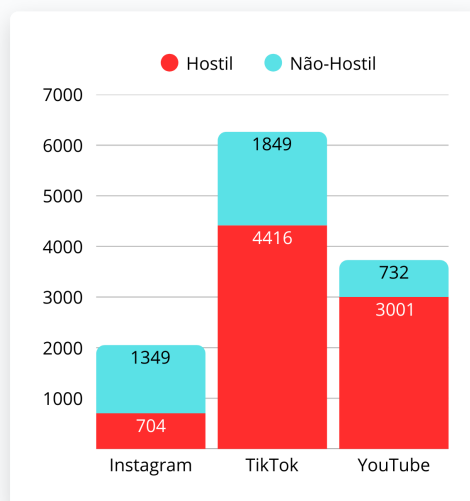


Gráfico 9



As diferenças percentuais, contudo, não representam níveis distintos de gravidade, mas sim **modos complementares de operação algorítmica**.

O TikTok concentrou a maior explosão de volume — **6.265 comentários em poucos dias**, sendo **mais de 70% hostis** —, enquanto o YouTube exibiu **a maior persistência e a pior proporção global: mais de 3.000 comentários hostis (80,4%)** mantidos ao longo de 15 meses, inclusive após o fim das postagens originais. O Instagram, embora tenha registrado o menor percentual (34,3%), não pode ser interpretado como um ambiente “protegido”: ali, a hostilidade se manifestou também por **outras vias discursivas e performáticas**, que não foram exploradas por este relatório: como **memes, remixes e vídeos de chacota** produzidos por terceiros, ou ainda **comentários transfóbicos nas postagens de figuras públicas que amplificaram o ataque**. Ou seja, o impacto no Instagram não foi necessariamente menor — apenas **mais difuso e esteticamente disfarçado**. Considerando a operação articulada entre plataformas, o efeito social e psicológico não se diferencia.

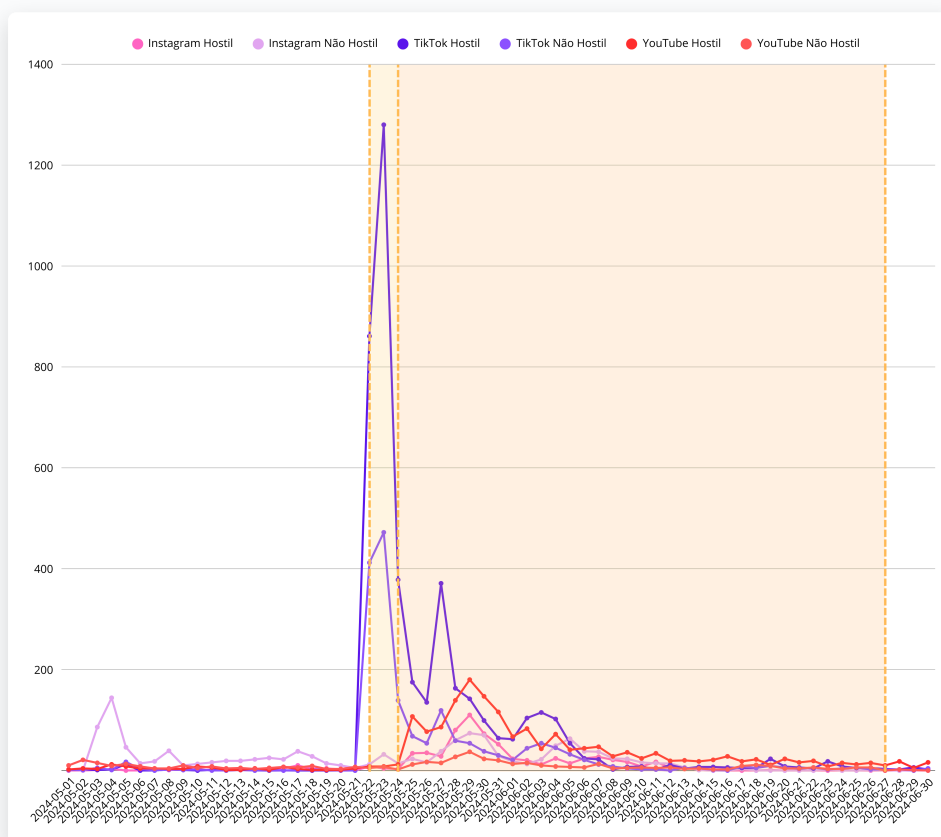


Gráfico 10

No plano temporal, o TikTok desempenha o papel de **gatilho inicial da onda**, com uma explosão antecipada em **22 de maio de 2024**, dois dias antes do vídeo de Nikolas Ferreira, o que indica **circulação prévia e orgânica** do conteúdo impulsionada pelo algoritmo de recomendação. Essa dinâmica, de alta concentração e curta duração, foi responsável por **gerar o primeiro pico emocional coletivo**, no qual conteúdos transfóbicos atingiram milhões de visualizações em questão de horas.

O YouTube, em contraste, age como **câmara de eco e arquivo ativo do ódio**. Após o início da onda, o percentual de hostilidade subiu de 38,2% para 83,7% e permaneceu estável acima de 80% por mais de um ano, atingindo 89,7% em 2025. Mesmo sem novas postagens, o canal continuou recebendo comentários transfóbicos, evidenciando **um comportamento autônomo de retroalimentação algorítmica**, em que a violência discursiva passa a ser um motor de engajamento permanente.

O Instagram, por sua vez, apresenta um **perfil reativo e dependente de atividade**, com aumento de 7,2% para 47,3% de comentários hostis/odiosos após o vídeo de Nikolas. Quando o perfil do podcast cessou as publicações, o volume despencou 97,3%, e o percentual caiu levemente (44,4%). Essa queda, contudo, **não indica resiliência da plataforma**, mas sim **a ausência de estímulos dentro do perfil analisado**. Fora dele, o conteúdo seguiu sendo explorado de forma transfóbica em remixagens, stories e postagens de influenciadores, configurando **um segundo ciclo de propagação simbólica** que extrapola a métrica de comentários. O Instagram, portanto, **não oferece uma resposta institucional mais eficaz** — apenas **uma aparência de contenção** que se deve à natureza fragmentada da análise de comentários, não à moderação real.

A complementaridade entre plataformas cria uma **cadeia funcional de difusão**. A circulação de conteúdos transfóbicos entre esses ambientes **estruturou uma narrativa de pânico moral multi-redes**, impulsionada por agentes políticos de extrema-direita e amplificada por arquiteturas algorítmicas que privilegiam emoção, conflito e engajamento.

Ao observar os dados de forma integrada, nota-se que a diferença entre as plataformas **não é de natureza moral, mas de arquitetura e tempo de resposta**. Nenhuma delas foi capaz de conter o ciclo de propagação. Em conjunto, criaram um ecossistema em que **a transfobia se torna viral, arquivável e compartilhável**, sustentando tanto a produção de capital simbólico por agentes extremistas quanto a

normalização da violência dirigida a pessoas trans e não binárias no espaço público digital.

## Conclusão e recomendações

O caso do *Entre Amigos* revela de forma contundente **como o discurso de ódio transfóbico se estrutura, se propaga e se mantém nas plataformas digitais** — não como evento isolado, mas como **um fenômeno sistêmico** sustentado por **arquiteturas tecnológicas que privilegiam o engajamento a qualquer custo**. A cronologia da onda demonstra um padrão recorrente: uma ação coordenada de atores políticos transfóbicos, impulsionada por algoritmos de recomendação que amplificam conteúdos de alto impacto emocional, seguida da **ineficácia das ferramentas de denúncia e da ausência de resposta institucional efetiva**.

Na **primeira fase**, de resposta rápida e triagem, evidenciou-se a **assimetria operacional** entre as vias de denúncia por discurso de ódio e por direitos autorais. Enquanto as infrações de *copyright* são tratadas com rastreabilidade, prazos e desfechos claros, as denúncias de discurso de ódio enfrentam **silêncio, opacidade e desresponsabilização**. Entre setembro e dezembro de 2024, foram registradas **98 denúncias**, com resultados que confirmam a discrepância estrutural: **58% de remoção em casos de direito autoral** contra apenas **8% em denúncias por discurso de ódio**. Em todas as plataformas, a via autoral foi ao menos **50 pontos percentuais mais eficaz**. O dado expõe uma política institucional de priorização econômica: plataformas agem com rigor para proteger a propriedade intelectual — interesse comercial — e negligenciam a proteção de pessoas LGBTQIA+, mesmo diante de ataques massivos.

A fase de análise de comentários reforçou empiricamente esse diagnóstico. A classificação manual (Instagram) e automatizada (TikTok e YouTube) de mais de 12 mil comentários revelou padrões alarmantes de hostilidade: 33,56% de hostilidade/ódio no Instagram, 70,5% de hostilidade no TikTok e 80,4% de hostilidade no YouTube. O dado, por si só, já seria alarmante; mas seu comportamento temporal e sua distribuição entre plataformas expõem um problema ainda mais profundo.

O TikTok gerou a explosão inicial de hostilidade — antecipando-se à postagem do deputado Nikolas Ferreira — e criou as condições para a fabricação de um factóide transfóbico de alta potência emocional, que alimentou a reação coordenada nas demais redes. O YouTube funcionou como repositório e câmara de eco, sustentando a hostilidade de forma persistente por mais de 15 meses, com mais de 80% de prevalência média de comentários hostis. O Instagram, embora com percentuais menores (34,3% de hostilidade/ódio), refletiu o impacto do ataque em escala relevante e serviu como superfície de reencenação simbólica, principalmente por meio de remixagens e comentários de chacota. Nenhuma das plataformas demonstrou diligência ativa na contenção do fenômeno.

O conjunto dos achados revela um problema de design institucional. As ferramentas internas de denúncia foram desenhadas para incidentes atomizados, e não para ataques coordenados. Essa arquitetura, aliada à ausência de auditoria pública sobre critérios de moderação, reforça o ciclo de impunidade: o ataque se multiplica, as vítimas se retraem, e o discurso de ódio permanece monetizado. Mesmo quando as vítimas acionam as plataformas diretamente, as respostas são automáticas, sem fundamentação verificável. Em alguns casos, como no YouTube, **100% das denúncias de ódio ficaram apenas com status "Recebida"**, sem qualquer desfecho.

Do ponto de vista jurídico, o caso ilustra **a pertinência da reinterpretação do art. 19 do Marco Civil da Internet** pelo Supremo Tribunal Federal. O entendimento de que plataformas devem exercer **diligência ativa em situações de circulação massiva de conteúdos gravíssimos** encontra aqui sua plena justificativa. O *Entre Amigos* não enfrentou apenas uma falha operacional, mas **uma negligência sistêmica** — a ausência de mecanismos que detectem, mitiguem e reparem campanhas coordenadas de violência de gênero mediadas por tecnologia.

Além disso, o episódio demonstra **a assimetria jurídica de proteção**: o direito autoral é reforçado por tratados internacionais e sanções comerciais (como a DMCA), enquanto os direitos humanos digitais carecem de *enforcement* equivalente. Essa desigualdade normativa é reproduzida pelas plataformas e molda sua governança de forma deliberada, priorizando o lucro e a estabilidade jurídica de suas operações sobre a proteção de grupos vulneráveis. O caso evidencia que o combate à transfobia e ao discurso de ódio não pode depender apenas de denúncias manuais ou de

medidas judiciais individuais, mas requer **responsabilidade estrutural, interoperabilidade de dados e auditoria contínua** das práticas de moderação.

Por fim, os resultados apontam para uma **economia política da visibilidade**, em que a violência se converte em produto. A monetização de vídeos e comentários com alto engajamento — inclusive os de ódio — cria incentivos perversos à sua perpetuação. Uma frase de um influencer conservador no Youtube resume esta plena consciência por parte dos criadores de conteúdo, sobre como as redes sociais recompensam o “hype” de conteúdos de ódio LGBTfóbico:

*“Na verdade, ninguém tá nem aí pra vocês. A única coisa que as pessoas veem nisso é uma palhaçada mesmo, as pessoas veem essas coisas que vocês falam, esses termos que vocês inventam, como uma palhaçada, a gente só ri, acha isso engraçado, usa vocês de hype, tchau, é só pra isso que vocês estão servindo no momento”*

Os algoritmos não são neutros: atuam como **curadores de relevância**, amplificando o que provoca reações extremas e silenciando conteúdos que geram reflexão. Ao responder positivamente a estratégias de mobilização política da extrema-direita, as plataformas **participam ativamente da construção de um ambiente digital hostil à diversidade e à democracia**. O caso evidencia que o combate ao discurso de ódio não pode depender apenas de denúncias manuais ou medidas judiciais posteriores, mas requer **responsabilidade estrutural, interoperabilidade de dados e auditoria contínua das práticas de moderação**.

## Recomendações:

---

**Transparência algorítmica obrigatória:** divulgação pública das métricas de recomendação, engajamento e moderação, com auditoria independente sobre a priorização de conteúdos que envolvem ataques a grupos vulneráveis.

**Mecanismos de detecção coletiva:** identificação automatizada e humana de surtos coordenados de discurso de ódio (picos anômalos de menções e comentários ofensivos).

**Reforma das interfaces de denúncia,** com caminhos simplificados para campanhas coletivas de assédio, sem exigir o registro manual de cada comentário ou postagem.



**Relatórios públicos trimestrais:** detalhamento de remoções e desempenho de moderação, segmentados por tipo de violência (racial, de gênero, LGBTfóbica etc.).

**Instâncias mistas de revisão:** com participação de sociedade civil, pesquisadores e grupos afetados, para avaliar casos graves e reincidentes.

**Reconhecimento jurídico da VGFT:** violência de gênero facilitada por tecnologia como categoria autônoma de dano, exigindo resposta proporcional à escala e à coordenação dos ataques

**Políticas de reparação simbólica e material:** desmonetização de conteúdos ofensivos e reinvestimento compulsório em educação digital, diversidade e segurança de grupos vulneráveis.

**Criação de mecanismos de interoperabilidade e resposta em rede:** para que plataformas troquem sinais de alerta sobre campanhas coordenadas, impedindo a migração e recirculação de conteúdos de ódio entre ambientes digitais.

Em síntese, o caso *Entre Amigos* demonstra que o problema não é apenas de moderação, mas de **governança da informação e responsabilização algorítmica**. A transfobia e o discurso de ódio observados são sintomas de uma estrutura que transforma a violência em valor de engajamento. Enfrentar esse modelo requer medidas regulatórias e culturais que reconheçam que o dano não está apenas nas palavras, mas nas **infraestruturas que as amplificam, monetizam e perpetuam**.

## Glossário

**Arquitetura de plataforma** — Estrutura técnica e normativa que define como uma plataforma organiza, prioriza e monetiza conteúdos, moldando o comportamento e a visibilidade dos usuários.

**Boyceta** — Termo que emerge das batalhas de rima de pessoas dissidentes de gênero no Largo de São Bento, em São Paulo, em 2018, da rima de Roberto Inácio, para nomear corpos e experiências de dissidência de gênero transmasculinas que não se encaixam na nomenclatura de homem trans. Junção das palavras "Boy" e "Buceta", representa pessoas transmasculinas que se apoderam de seus corpos e não recusam suas genitálias.

**Cadeia de custódia digital** — Registro documentado e contínuo das etapas de coleta, armazenamento e preservação de evidências digitais, garantindo sua integridade jurídica.

**Contradiscorso** — Expressão ou conteúdo produzido como forma de resistência e resposta a narrativas de ódio e desinformação.

**Cisheteronormatividade** — Regime cultural que considera a cisgeneridade e a heterossexualidade como padrões naturais e desejáveis, marginalizando outras identidades e orientações.

**Desenho de produto** — Concepção técnica e funcional de uma plataforma ou ferramenta digital, incluindo a definição de fluxos, arquitetura da informação e regras de interação. O desenho de produto influencia diretamente como usuários se comportam e como a moderação de conteúdo ocorre.

**Devida diligência** — Princípio jurídico segundo o qual atores públicos e privados têm o dever de prevenir, investigar e remediar danos previsíveis. No contexto das plataformas digitais, implica a obrigação de agir de forma proativa diante de conteúdos gravemente ilícitos, conforme a reinterpretação do Supremo Tribunal Federal ao artigo 19 do Marco Civil da Internet.

**Discurso de ódio** — Expressão ou comunicação que incita, propaga ou legitima hostilidade contra pessoas ou grupos com base em características como gênero, raça, orientação sexual, identidade de gênero ou religião.

**Direito autoral / Copyright** — Conjunto de normas que protege a autoria e o uso de obras intelectuais. No relatório, serve de contraponto para demonstrar a maior eficiência das plataformas na proteção de interesses econômicos em relação aos direitos humanos.

**Economia da visibilidade** — Lógica de mercado das plataformas digitais em que a atenção e o engajamento se convertem em valor econômico, inclusive quando derivados de conteúdo violento ou discriminatório.

**Enforcement** — Execução prática e efetiva de normas ou políticas, especialmente em ambientes digitais. No relatório, o termo descreve o grau de aplicação real das regras

de uso e das diretrizes de comunidade pelas plataformas — distinguindo o que é formalmente previsto do que é realmente cumprido.

**Fair use** — Doutrina jurídica norte-americana que permite uso limitado de obras protegidas sem autorização, mas que não se aplica ao ordenamento jurídico brasileiro.

**Hostilidade** — Termo amplo que designa qualquer forma de agressão, desrespeito, ataque ou comportamento negativo dirigido a pessoas ou grupos. No relatório, é utilizado em dois contextos: (1) No Instagram, designa especificamente comentários de severidade 1; (2) No TikTok e YouTube, onde a análise binária por IA não permite distinguir níveis de severidade, "hostilidade" é o termo utilizado para toda a categoria de comportamentos hostis identificados pelo modelo. Diferencia-se de "ódio", que é uma subcategoria específica de hostilidade referente a severidade 2-3. Ver escala de severidade na seção metodológica.

**Litigância estratégica** — Uso deliberado do sistema judicial para provocar mudanças estruturais, gerar precedentes e ampliar a proteção de direitos humanos.

**Marco Civil da Internet (art. 19)** — Norma brasileira que regula a responsabilidade das plataformas por conteúdos de terceiros, reinterpretada pelo STF para exigir ação proativa em casos de discurso de ódio.

**Moderação reativa** — Prática em que plataformas digitais só agem para remover conteúdos após denúncia, sem mecanismos preventivos ou proativos de controle.

**Misgendering / Deadnaming** — Práticas de violência simbólica que consistem em negar ou desrespeitar a identidade de gênero de uma pessoa trans ao usar o nome ou o pronome de antes da transição.

**Pânico moral** — Reação social de medo e indignação desproporcional frente a grupos ou práticas marginalizadas, instrumentalizada politicamente para justificar repressão.

**Raspagem de dados (data scraping)** — Técnica de coleta automatizada de informações públicas em plataformas digitais para análise quantitativa e qualitativa.

**React (vídeo de reação)** — Formato de conteúdo em que uma pessoa grava suas expressões, comentários e respostas ao assistir a outro vídeo, música ou publicação. Popular nas redes sociais, este estilo enfatiza a interação afetiva e opinativa do público com o material original, funcionando como uma forma de engajamento,

comentário social ou reforço de tendências algorítmicas. No contexto do relatório, o termo designa vídeos de reação criados por influenciadores para amplificar ataques ou reforçar discursos transfóbicos a partir de conteúdos do Entre Amigos.

**Social listening** — Processo de monitoramento sistemático de menções, palavras-chave e interações em redes sociais para identificar tendências, discursos e comportamentos. No relatório, refere-se à análise automatizada feita com a ferramenta Brand24 para rastrear o uso dos termos “boyceta” e “Entre Amigos”.

**Violência de gênero facilitada por tecnologia (VGFT)** — Forma de violência baseada em gênero mediada por meios digitais, que inclui assédio, exposição não consentida, perseguição e campanhas coordenadas de ódio.

## Referências

BRASIL. Superior Tribunal de Justiça. Agravo Regimental no Recurso Ordinário em Habeas Corpus n. 184.003/SP. Relatora: Ministra Daniela Teixeira. Brasília, DF, 10 de dezembro de 2024. Diário da Justiça Eletrônico, 26 dez. 2024.

BUTLER, Judith. *Problemas de gênero: feminismo e subversão da identidade*. Rio de Janeiro: Civilização Brasileira, 2003.

FOUCAULT, Michel. *História da sexualidade 1: a vontade de saber*. 13. ed. Rio de Janeiro: Graal, 1999a.

FOUCAULT, Michel. *Vigiar e punir: nascimento da prisão*. 20. ed. Petrópolis: Vozes, 1999b.

HARAWAY, Donna. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective". *Feminist Studies*, v. 14, n. 3, 1988, p. 575-599.

PRECIADO, Paul B. "Multidões queer: notas para uma política dos 'anormais'". *Revista Estudos Feministas*, Florianópolis, v. 19, n. 1, p. 11-20, 2011.

PRECIADO, Paul B. *Testo Junkie: Sexo, drogas e biopolítica na era farmacopornográfica*. Zahar, 2023.

SHECAIRA, Fernando Muniz. *A cifra arco-íris: acesso à justiça, população LGBTI+ e interseccionalidades*. Tese (Doutorado em Direito) — Faculdade de Direito, Universidade de São Paulo, São Paulo, 2025, 410 f., orient. Susana Henriques da Costa.

## Créditos

Código Não Binário

[codigonaobinario.com.br](http://codigonaobinario.com.br)

[contato@codigonaobinario.org](mailto:contato@codigonaobinario.org)

+55 11 98109-6572

### Licença:

CC BY-SA 4.0

Creative Commons Attribution-ShareAlike 4.0 International