

Analizando a Anatomia do Discurso de Ódio nas Interações das Comunidades da Plataforma de Jogos Steam

Lucas Martins Souza
Universidade Federal de Viçosa
Florestal - MG, Brasil
lucas.souza59@ufv.br

Paula T. M. Gibrim
Universidade Federal de Viçosa
Viçosa - MG, Brasil
paula.gibrim@ufv.br

Marcus V. G. Ribeiro
Universidade Federal de Viçosa
Viçosa - MG, Brasil
marcus.guerra@ufv.br

Daniel M. Barbosa
Universidade Federal de Viçosa
Florestal - MG, Brasil
danielmendes@ufv.br

Julio C. S. Reis
Universidade Federal de Viçosa
Viçosa - MG, Brasil
jreis@ufv.br

Philipe de Freitas Melo
Universidade Federal de Viçosa
Florestal - MG, Brasil
philipe.freitas@ufv.br

ABSTRACT

This study examines the anatomy of hate speech on Steam, one of the world's largest game distribution and social interaction platforms. Using a dataset of 34.300.919 comments and a lexicon of 1,298 terms, the investigation identifies where these manifestations occur, as well as their targets and themes. The results reveal a systemic presence of hate, with particular emphasis on attacks based on ethnicity, disability, social class, and sexual orientation. The analysis shows that toxicity is influenced by community context and game tags. Finally, topic modeling highlights how hate is often disguised as technical criticism, reinforcing the need for moderation mechanisms that are more sensitive to context.

KEYWORDS

Hate speech; Toxicity; Steam; Topic modeling; Content moderation

1 INTRODUÇÃO

Lançada em 2003 pela Valve Corporation, a Steam consolidou-se como a principal plataforma de distribuição digital de jogos para computadores pessoais, reunindo atualmente mais de 132 milhões de usuários cadastrados e registrando um pico de 42,04 milhões de usuários simultâneos no início de 2026 [2]. Seu catálogo supera 230 mil títulos, abrangendo desde produções independentes (*indie games*) até títulos de grande orçamento [3]. Em termos de participação de mercado, a plataforma concentra aproximadamente 75% do segmento de jogos digitais para PC, respondendo por cerca de 18% das vendas totais da categoria, excluindo-se conteúdos adicionais (DLCs) e microtransações [27]. Esses indicadores posicionam a Steam não apenas como o maior ecossistema de distribuição de jogos digitais do mundo, mas também como um objeto de estudo de crescente relevância para a análise de comportamentos de usuários em ambientes digitais em larga escala.

Muito além de uma plataforma comercial, a Steam apresenta características que a alinham à definição de rede social proposta por Boyd e Ellison [6]: um serviço online que permite aos indivíduos construir perfis públicos ou semipúblicos, articular listas de conexões com outros usuários e transitar por essas redes dentro do próprio sistema. A plataforma satisfaz plenamente esses critérios, uma vez que seus usuários mantêm perfis detalhados com histórico de jogos e conquistas, estabelecem vínculos sociais por meio de listas de amigos e engajam em grupos, fóruns e sistemas de avaliação que conferem visibilidade pública a essas interações [36]. Dessa forma, a Steam ocupa uma posição singular no ecossistema digital, operando simultaneamente como vasto mercado consumidor e espaço de socialização em larga escala.

Contudo, como em redes sociais tradicionais, essa dimensão social na plataforma abre espaço para manifestações de ódio. A literatura já documenta amplamente como plataformas de interação online favorecem a disseminação de discurso de ódio, comportamentos disruptivos e dinâmicas de toxicidade [12, 23]. Na Steam, esse fenômeno assume uma forma particular: a plataforma agrega uma comunidade orientada por um interesse comum, o jogo, o que pode tanto intensificar os laços sociais quanto acirrar conflitos identitários e culturais [20, 33]. Grupos historicamente marginalizados, como mulheres e minorias étnicas, relatam experiências recorrentes de hostilidade nesse ambiente [13] – fenômeno que ganhou visibilidade internacional com o *Gamergate*, campanha de assédio organizado que expôs as tensões identitárias latentes da comunidade *gamer* [26]. Além disso, episódios como o do jogo *Total War: Rome II*, que recebeu uma onda de análises misóginas após a inclusão de generais mulheres no sistema [1, 8, 35] reforçam que o discurso de ódio nesse contexto não é episódico, mas estrutural.

Embora os estudos sobre discurso de ódio online tenham avançado consideravelmente na última década, investigações em larga escala voltadas especificamente para plataformas de jogos digitais permanecem escassas [15]. Estudos geralmente focam em toxicidade e discurso de ódio gerais, não deixando claro o processo de vitimização e vulnerabilidade de grupos específicos. A normalização dessa toxicidade expõe crianças

e adolescentes, usuários em fase de desenvolvimento, à interações abusivas sem a devida mediação [20]. É justamente a urgência em mitigar essa vulnerabilidade que impulsionou recentemente, no Brasil, a aprovação do Estatuto Digital da Criança e do Adolescente (ECA Digital), cujo foco recai na proteção desse público contra conteúdos nocivos em ambientes virtuais. Isso contrasta com a expressividade do setor: o mercado global de jogos digitais registra crescimento contínuo [28], e plataformas como a Steam concentram volumes massivos de interação textual cotidiana [31, 37]. Compreender como o discurso de ódio se manifesta nesse ambiente torna-se, portanto, uma necessidade tanto científica quanto social.

Diante desse cenário, este trabalho tem como objetivo analisar e caracterizar a presença dos diferentes tipos de discurso de ódio presente nos comentários de avaliação da Steam, investigando padrões de ocorrência desse tipo de conteúdo, os contextos em que se concentra e suas formas mais recorrentes. Para tanto, foram coletados quase 90 milhões de comentários de mais de 150 mil jogos. Realizamos então a detecção de toxicidade automatizada em todo nosso conjunto de dados, em conjunto com a identificação de termos de discurso de ódio ancorada na *Hatebase*¹ – repositório multilíngue e estruturado baseado em *crowdsourcing* cuja taxonomia fornece o referencial categórico adotado ao longo do estudo.

Por “anatomia”, em linha com a análise de composição de [17], entendemos a decomposição do discurso de ódio em suas dimensões constituintes, onde ocorre, contra quem se dirige e de que forma se expressa, aqui aplicada não a bases de dados rotuladas, mas à manifestação do fenômeno na própria Steam. Dessa forma, visamos responder as seguintes questões de pesquisa: (QP1) Onde é prevalente o discurso de ódio nos comentários de análise dos jogos da Steam? (QP2) Quais são os alvos do ódio propagado nesse ecossistema? (QP3) Quais assuntos são frequentemente encontrados nesses comentários?

Os resultados demonstram que a manifestação do discurso de ódio na plataforma Steam é fortemente condicionada pelo contexto e pelas características da comunidade associada a cada jogo, evidenciando a heterogeneidade dos cenários e o desafio inerente à sua compreensão. A análise revela, ademais, que tanto a dinâmica quanto os alvos dos ataques variam de forma expressiva conforme as *tags* dos títulos avaliados, abrangendo desde hostilidades direcionadas às equipes de desenvolvimento dos jogos até ofensas diretas e condutas de degradação entre jogadores em títulos competitivos de esportes eletrônicos (*e-sports*). Por fim, a modelagem de tópicos evidencia a perpetuação sistêmica de discursos racistas, sexistas e capacitistas, frequentemente dissimulados em meio a críticas técnicas ou xingamentos de caráter genérico o que reforça a complexidade dos mecanismos de moderação nesses espaços e aponta para a necessidade de abordagens sensíveis ao contexto.

O restante deste artigo está organizado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados; a Seção 3 descreve a metodologia adotada; as Seções 4, 5 e 6 discutem

os resultados obtidos; e a Seção 7 traz as conclusões e perspectivas de trabalhos futuros.

2 TRABALHOS RELACIONADOS

Diante da relevância e gravidade do fenômeno, o discurso de ódio online tem atraído atenção crescente da comunidade científica. Pesquisadores discutem os desafios inerentes ao fenômeno [22], à sua identificação [19] e propõem recursos para auxiliar soluções automatizadas a fim de mitigar esse cenário [12, 30]. Grande parte desses esforços, no entanto, concentra-se em redes sociais tradicionais, onde a ambiguidade e a subjetividade da linguagem já impõem dificuldades contínuas à detecção automática.

Um dos principais desafios metodológicos nesse campo reside na escolha, na qualidade e nos vieses das bases de dados utilizadas para categorizar o discurso de ódio. Avaliando essa questão, [17] analisaram a composição e a taxonomia de seis bases de discurso de ódio majoritariamente extraídas do Twitter/X, amplamente empregadas na literatura, demonstrando como a distribuição desbalanceada de alvos de ataques afeta severamente a capacidade de generalização dos modelos classificadores. De forma análoga, [25] investigaram repositórios em português brasileiro voltados à detecção de discurso de ódio e toxicidade, evidenciando que a diversidade lexical e as estratégias de rotulagem adotadas na curadoria dos dados são determinantes para a robustez dos sistemas de classificação. Esses resultados reforçam a importância de investigar novas fontes de dados que ampliem a diversidade de domínios analisados, fugindo das plataformas e dos léxicos convencionalmente adotados na literatura.

Em paralelo, o ecossistema dos jogos digitais se consolidou como objeto de interesse crescente da academia. Diversos trabalhos analisam a influência desses jogos no público infantojuvenil e sua importância no desenvolvimento cognitivo e na formação social dos indivíduos [7, 10, 14, 32]. Contudo, apesar de seus aspectos formativos e de entretenimento, as comunidades construídas nesse ecossistema são frequentemente marcadas por manifestações odiosas e comportamentos disruptivos.

A literatura tem refletido essa preocupação de forma crescente. Na revisão de [15], foram levantados 49 estudos sobre discurso de ódio e toxicidade em jogos digitais. O trabalho revelou que a literatura foca primariamente na dinâmica interna de títulos específicos como DOTA 2 [24] e *League of Legends* [5], com foco maior na comunicação *in-game* (no software em si). Embora estudos recentes tenham começado a explorar a detecção de toxicidade no contexto mais amplo de plataformas voltadas a jogos [9, 29] como na Twitch [23] e no Discord [18], ainda há um baixo volume de análises transversais e longitudinais focadas na infraestrutura social que hospeda e distribui esses jogos. Indo além da análise de títulos ou plataformas isolados, [34] investigaram o *review bombing* no Metacritic, demonstrando como sistemas de avaliação funcionam como mecanismos de pressão social e espelhos de tensões comunitárias. Ainda assim, a Steam permanece

¹<https://hatebase.org>

subexplorada, mesmo com sua influência global e relevância mercadológica.

Este trabalho se posiciona, portanto, na interseção entre a taxonomia em larga escala de discursos de ódio e o estudo transversal de ecossistemas de jogos. Partindo da metodologia e do léxico da Hatebase, já consolidados por [17], este estudo escala a investigação para milhões de comentários de múltiplas comunidades dentro da Steam. Ao fazer isso, o artigo responde à lacuna de análises multicontextuais na área, deslocando o foco da predição algorítmica para o entendimento semântico massivo do fenômeno. O resultado é uma primeira anatomia do discurso de ódio na loja da Steam, detalhando não apenas os alvos primários dessa toxicidade, mas também os contextos, nichos e as tags onde essas manifestações ganham espaço.

3 METODOLOGIA

Esta seção detalha a metodologia para a construção e análise da base de dados. As etapas compreendem a coleta massiva de comentários na Steam, seguida pela identificação de termos de ódio e levantamento do grau de toxicidade. A arquitetura geral desse pipeline, detalhando os passos de filtragem até a consolidação do conjunto final de comentários odiosos, é ilustrada na Figura 1.

3.1 Coleta de Dados da Steam

A coleta de dados de comentários em avaliações de jogos foi realizada por meio da API oficial da Steam² e estruturada em duas etapas principais. Na primeira, obtivemos a lista completa de 157.825 jogos disponíveis na plataforma, juntamente com suas respectivas informações – e.g., nome, número de comentários, gêneros e *tags*. Enquanto o gênero se refere à categoria do jogo, as *tags* da Steam são marcadores populares, atribuídos pela comunidade, para descrever características gerais de jogabilidade, conteúdo ou público-alvo. Em seguida, foram removidos os jogos que não possuíam comentários, resultando em um conjunto final de 83.420 jogos.

Na segunda etapa, foram coletados todos os comentários associados a esses jogos presentes em suas páginas na plataforma, totalizando 89.986.187 comentários em idiomas diversos, cada um contendo informações como o indicador de recomendação (positivo ou negativo), data de publicação e texto da análise; a coleta foi realizada entre setembro de 2024 e abril de 2025. Cabe destacar que, em contextos de processamento de linguagem natural, o idioma das mensagens exerce forte influência sobre as análises realizadas [12]. Ao examinar os idiomas mais frequentes no conjunto de dados coletado, temos uma predominância do inglês, com 34.300.919 comentários ($\approx 38\%$ do total), seguido pelo chinês, com 11.236.495 comentários ($\approx 12.5\%$) e português com 6.448.756 ($\approx 7.16\%$). Em razão dessa distribuição, o presente trabalho restringiu sua análise ao conjunto de comentários de análise redigidos em inglês.

²<https://steamcommunity.com/dev>

Tabela 1: Distribuição dos termos na Hatebase por tipo de discurso, incluindo a volumetria e exemplos do vocabulário.

Atributo	Tradução	Total	Exemplos
<i>ethnicity</i>	etnia	950	negro, niggerdick, Christ killer, monkey
<i>nationality</i>	nacionalidade	299	ice nigger, surrender monkey, shanty Irish, leprechaun
<i>class</i>	classe social	112	shit kicker, sheepfucker, bog trotter, scuffer
<i>religion</i>	religião	102	muslimic, Jihadi, holo-hoax, jewtarded
<i>gender</i>	gênero	45	bitches, gender nigger, fudgepackers, slut
<i>sexual orientation</i>	orientação sexual	36	faggot, gaylord, cocksuckers, fruit
<i>disability</i>	deficiência	23	idiot, mongols, retarded, cripples
Termos únicos		1298	

3.2 Léxico da Hatebase

No contexto de mineração de texto, um léxico consiste em um conjunto de palavras ou expressões pré-definidas que serve como base para identificar a presença de determinados padrões linguísticos em sentenças [4]. Diante disso, foi selecionada a Hatebase, uma base de dados colaborativa multilíngue com 1298 expressões que funciona como léxico de terminologias recorrentes em discursos de ódio. Cada termo catalogado da *Hatebase*, ainda, é organizado de acordo com o grupo-alvo ao qual se refere (e.g., a categoria *gênero* agrupa termos de ataques motivados por identidade de gênero).

A Tabela 1 apresenta 7 categorias do léxico, suas respectivas traduções para o português, o volume de termos associados a cada uma e exemplos de vocábulos (em inglês) que as compõem. Cabe ressaltar que a categorização adotada pela Hatebase não é exclusiva, i.e., um mesmo termo pode pertencer simultaneamente a mais de uma categoria. Assim, a soma dos totais por categoria é superior ao número de termos únicos.

3.3 Definição de Mensagens Odiosas

Para guiar a definição de ódio adotada por este trabalho, foram utilizados dois critérios complementares: a ocorrência de termos potencialmente odiosos e um alto nível de toxicidade no texto definido pela pontuação atribuída por uma API. Dessa forma, definimos **comentários com presença de discurso de ódio**, ou **comentário odioso**, como todo comentário cujo texto atenda às duas métricas. Primeiramente, verificamos a ocorrência de termos do léxico de ódio nos comentários coletados, adequando o primeiro critério. Assim, filtramos os comentários de modo a preservar apenas aqueles que apresentassem ao menos um termo pertencente ao léxico. Em seguida, realizamos um pós-processamento por meio da *Perspective API*³ em toda a base de dados para automaticamente atribuir um valor de toxicidade de cada comentário. Essa ferramenta é um modelo de *deep learning* treinado que fornece uma pontuação no intervalo 0,1 indicando a probabilidade de o

³<https://www.perspectiveapi.com/>



Figura 1: Fluxograma com a arquitetura geral da metodologia do trabalho.

conteúdo ser tóxico, além de identificar o idioma do texto. Cabe destacar que a plataforma Steam emprega mecanismos automáticos de moderação que censuram palavrões e termos ofensivos, substituindo-os pelo símbolo ♥. Entretanto, nossa abordagem contorna esse problema e evidencia que mesmo com a moderação da plataforma conseguimos identificar uma quantidade substancial de comentários odiosos por meio dos termos da HateBase e *pontuação* da Perspective.

Dessa maneira, foram identificados 1.318.897 comentários que atenderam ao critério de apresentar ao menos 1 termo do léxico, evidenciando a expressiva ocorrência desse tipo de conteúdo na plataforma. Para avaliar o segundo requisito, adotamos o limiar de 0,8 na pontuação de toxicidade gerada pela API, valor consistente com o empregado na literatura [11, 21], resultando em 274.978 comentários classificados como tóxicos. A interseção dos dois critérios, presença de termo do léxico de ódio e pontuação de toxicidade superior a 0,8, compõe o conjunto de dados final, constituído por 13.811 comentários odiosos.

4 DISCURSO DE ÓDIO NA PLATAFORMA

Com a base consolidada, iniciamos a análise investigando a relação entre o vocabulário das avaliações e a toxicidade detectada. A Tabela 2 apresenta a distribuição dos termos por categoria da HateBase e a proporção que efetivamente resultou na classificação de discurso de ódio, refletindo a representatividade das categorias sendo que os tipos que aparecem em menor volume no léxico, mas em grande quantidade nos comentários apresentam uma razão alta e razão baixa caso contrário. O ponto central dessa análise reside na disparidade entre volume e gravidade: a mera presença de um termo do léxico resulta em uma taxa de toxicidade reduzida (1,05%), o que indica que grande parte desses termos ocorre em contextos não-ofensivos.

A categoria *etnia* domina as ocorrências absolutas (mais de um milhão de comentários) e concentra o maior número absoluto de comentários odiosos (6.034). Contudo, sua razão de toxicidade é de apenas 0,59%, indicando que seus termos se diluem em usos majoritariamente não-ofensivos ou ressignificados pela comunidade. Comportamento análogo é observado em *nacionalidade* (0,15%) e *religião* (0,09%). Em contrapartida, as categorias de menor volume absoluto concentram as maiores taxas de ódio: *deficiência* destaca-se com razão de 20,00%, seguida por *orientação sexual* (7,75%) e *classe social* (3,09%). Esse padrão evidencia que o vocabulário dessas categorias carrega um estigma mais acentuado

e é predominantemente empregado de forma pejorativa, independentemente do contexto.

Esse padrão é corroborado pela Função de Distribuição Acumulada (CDF) de toxicidade apresentada na Figura 2. A curva de referência, composta por comentários sem termos da HateBase, confirma que uma parcela mínima dessas avaliações atinge o limiar de toxicidade estabelecido. A presença de qualquer termo do léxico já desloca as curvas para a direita, indicando aumento do potencial ofensivo. Ainda assim, a análise reforça que volume não implica gravidade: a curva de *etnia*, embora volumosa, indica que menos de 1% dos seus comentários atingem um grau de toxicidade elevado. Em contraste, as curvas de *deficiência* e *orientação sexual* deslocam-se acentuadamente para a direita, com 20% e 8% dos seus comentários ultrapassando o limiar de discurso de ódio, respectivamente. Esses resultados sugerem que termos de alta frequência são frequentemente incorporados ao jargão das comunidades de jogos, ao passo que categorias de menor volume tendem a ser mobilizadas como ataques verbais diretos. Diante disso, abordagens metodológicas puramente generalistas mostram-se insuficientes para capturar as nuances inerentes a esse domínio. Como a Steam abriga públicos heterogêneos, as dinâmicas de interação divergem de forma significativa entre comunidades. Para preservar essas particularidades, as seções seguintes investigam o discurso de ódio com base na granularidade das *tags* atribuídas pelas comunidades (e.g., *LGBTQ+*, *Competitive*), utilizando exclusivamente o conjunto de 13.811 comentários odiosos consolidados.

Para compreender a evolução temporal do fenômeno, a Figura 3 exibe a distribuição anual do volume de comentários odiosos na plataforma. Observa-se um crescimento acentuado a partir de 2019, culminando em um pico superior a 2.500 ocorrências em 2023. Embora essa tendência reflita, em parte, o próprio crescimento da base de usuários da Steam, ela expõe a dificuldade crescente da plataforma em escalar seus mecanismos de moderação na mesma proporção. A distribuição por categoria revela que o fenômeno não se restringe a anomalias ou eventos isolados, mas configura uma presença sistemática ao longo do período analisado.

5 ÓDIO NOS JOGOS E TAGS

Nesta seção, investigamos os contextos da plataforma Steam em que o discurso de ódio ocorre com maior frequência (QP1),

Tabela 2: Ocorrências de ódio por categorias da HateBase.

Atributo	Coment. com algum termo	Coment. odiosos	Razão (%)
etnia	1.004.194	6.034	0,59
deficiência	69.047	4.896	20,00
classe social	180.840	3.688	3,09
orientação sexual	98.438	2.972	7,75
nacionalidade	247.646	490	0,15
gênero	57.168	481	1,00
religião	60.302	103	0,09
Coment. únicos	1.318.897	13.811	1,05

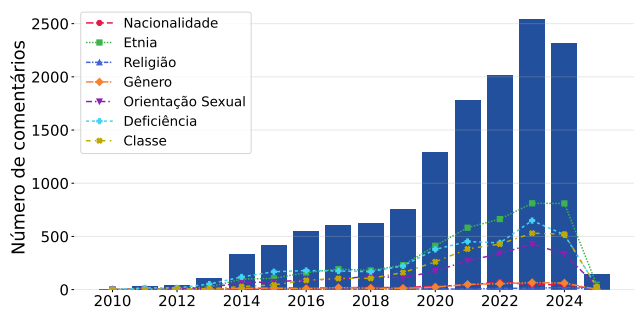


Figura 3: Volume de comentários de ódio ao longo dos anos.

examinando em que medida jogos e *tags* específicos estão associados a uma maior incidência desse fenômeno e as suas categorias predominantes. Para tanto, comparamos a proporção de comentários odiosos em relação ao total de comentários por jogo e por *tag*, permitindo identificar os contextos em que esse tipo de discurso se manifesta com maior intensidade.

5.1 Jogos mais Odiosos

A partir das definições anteriores, foram identificados os jogos e *tags* com maior proporção de comentários odiosos em relação ao total, sendo apresentados os 10 mais representativos de cada conjunto. Para assegurar representatividade estatística, a análise foi restrita a jogos e *tags* com volume mínimo de comentários correspondente ao percentil 75 de cada conjunto, totalizando 1.718 e 157.855 comentários, respectivamente. Cabe ressaltar que, embora as proporções observadas possam parecer reduzidas em termos absolutos, esses comentários correspondem a mensagens que lograram contornar os mecanismos de moderação da plataforma, permanecendo visíveis ao conjunto de usuários.

As Figuras 4a e 4b apresentam os jogos e *tags* com maior proporção de discurso de ódio. Entre os jogos, *Maize*, apesar de não apresentar o maior volume absoluto de comentários odiosos ocupa a primeira posição em termos relativos, destacando-se por concentrar ocorrências de uma única categoria de ódio, relacionada à deficiência. Nos comentários associados a esse título, observa-se a recorrência do termo

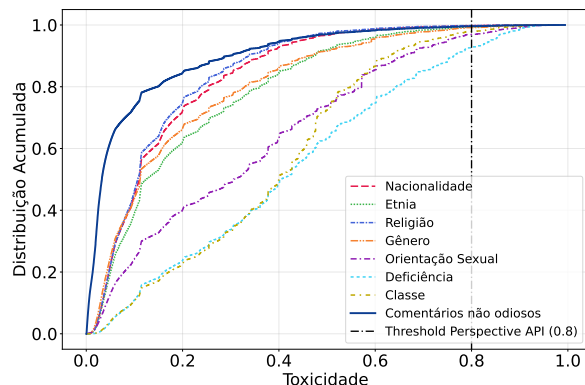


Figura 2: CDF toxicidade por categoria da HateBase.

stupid, frequentemente empregado com intenção humorística para descrever o jogo e seus personagens – uso que, ainda que contextualizado, não deixa de configurar discurso odioso, conforme o exemplo:

“Don’t be stupid, *idiot*. Vladdy demands you play this American garbage game.” [Comentário coletado do jogo *Maize*.]

Observa-se ainda a presença de termos associados ao discurso de ódio relacionado à orientação sexual em jogos como *Amorous*, *FPS Chess* e *Radiator 2*, com destaque para o termo *gay*, frequentemente empregado de forma pejorativa, como ilustra o exemplo:

“SHINY KNIGJHT IS *GAY* NERF PLSSS IOM GONNA KILL MYSLEF” [Comentário coletado do jogo *FPS Chess*.]

Em relação às *tags*, observa-se maior diversidade de categorias de ódio em comparação aos jogos, com destaque para *etnia*, *orientação sexual*, *classe* e *deficiência*. Algumas *tags* tendem a coocorrer nos mesmos jogos, como *Tabletop*, *Hidden Object* e *Board Game*, que agrupam predominantemente jogos de mesa, frequentemente no modo multijogador. Nesses contextos, o discurso de ódio é variado, mas tende a ser direcionado aos desenvolvedores, com uso recorrente de termos como *retarded*, *trash* e *gay*. Em contraste, a *tag eSports*, embora ocupe apenas a 9ª posição em proporção, registra o maior volume absoluto de comentários odiosos (794 ocorrências), com uma mudança expressiva no alvo do discurso: os ataques deixam de se dirigir aos desenvolvedores e passam a ser direcionados a outros jogadores, como ilustra o exemplo a seguir:

“makes me want to kill myself also makes me want to hate darker *coloured* players because they are just worse” [Comentário coletado do jogo *EA SPORTS FC™ 25*.]

5.2 Anatomia dos Tipos de Discurso de Ódio na Steam

Um segundo eixo de análise consistiu em identificar em quais contextos cada categoria de discurso de ódio presente na Hatebase se manifesta de forma mais expressiva, trazendo

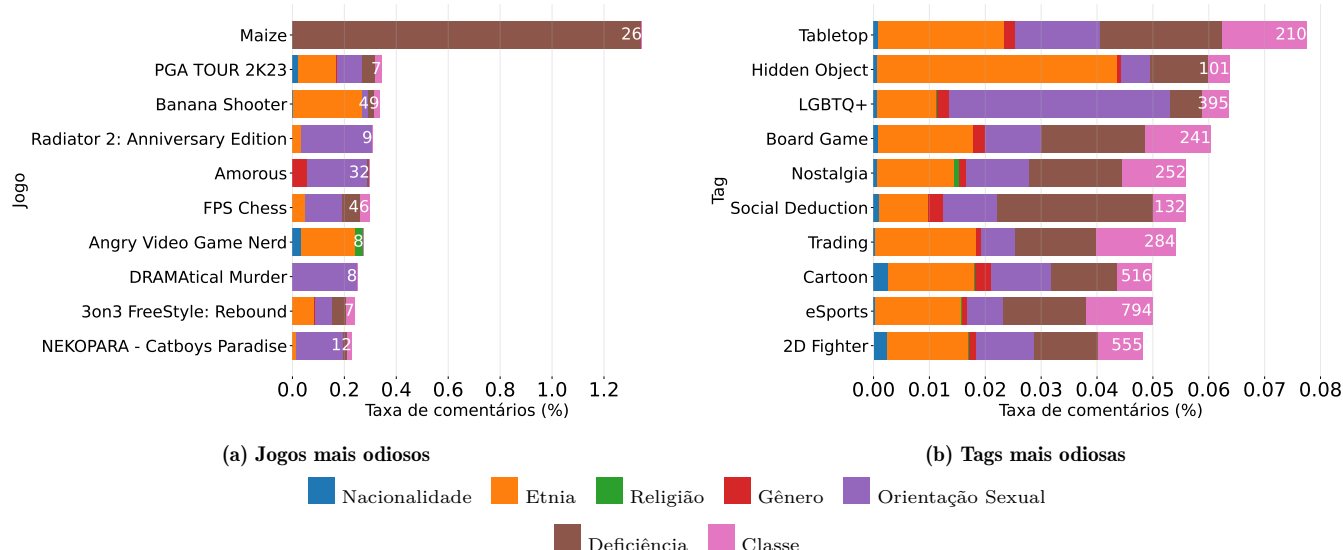


Figura 4: Jogos e tags com maior proporção (%) de comentários odiosos.

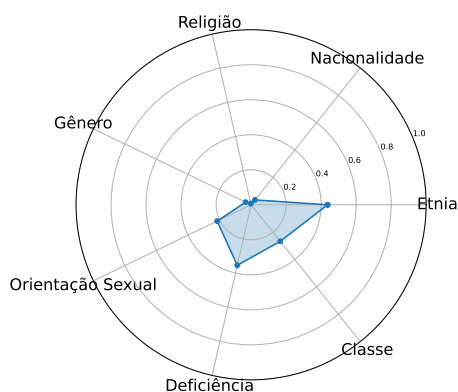


Figura 5: Tipologia dos comentários de ódio em geral.

a análise dos alvos do ódio (QP2). Para tanto, examinou-se inicialmente a distribuição global dos tipos de ódio nos comentários odiosos, apresentada na Figura 5. Os resultados indicam uma predominância de comentários relacionados à etnia, seguidos por deficiência, classe e orientação sexual, ao passo que nacionalidade, religião e gênero apresentam expressividade reduzida.

Em seguida, analisou-se exclusivamente os comentários odiosos associados a cada tag, calculando-se a proporção de cada tipo de ódio para identificar quais tags concentram maior incidência de cada categoria. Para assegurar representatividade estatística, foram consideradas apenas as tags correspondentes aos 50% com maior volume de comentários odiosos, limiar que demarca um aumento expressivo na densidade desse tipo de discurso. A distribuição resultante é apresentada na Figura 6, que exibe, para cada tipo de ódio, a tag com maior proporção de ocorrências daquela categoria.

A análise da Figura 6 revela que, embora determinadas tags concentrem a maior proporção de certo tipo de ódio, esse tipo nem sempre é o predominante em distribuição de

categorias de ódio. A tag *Demons*, por exemplo, lidera em comentários de ódio relacionados à nacionalidade, mas sua distribuição é dominada por etnia e deficiência (Figura 6a). De modo geral, corroborando o padrão global da Figura 5, etnia constitui a categoria predominante em praticamente todas as tags, com nacionalidade, religião e gênero mantendo baixa expressividade. Ainda assim, observam-se distribuições distintas: na tag *Trading*, a proporção de ódio relacionado à classe aproxima-se da de etnia; na tag *Turn Based Tactics*, os comentários referentes à deficiência superam os de etnia; e na tag *LGBTQ+*, embora etnia permaneça presente, os discursos relacionados à orientação sexual constituem a maioria significativa dos comentários odiosos.

“♥♥♥♥ you Ronnie ♥♥♥♥ ass scammer ♥♥♥♥ ass game deserves to be thrown in a trash can” [Comentário coletado da tag *Trading*.]

“GACHA. Roach. Bus. Idiot game for idiot players with idiot review comments.” [Comentário coletado da tag *Turn Based Tactics*.]

“I HATE THIS GAME THEIR IS GAY PEOPLE” [Comentário coletado da tag *LGBTQ+*.]

6 ANÁLISE TEMÁTICA DO DISCURSO DE ÓDIO NA STEAM

Após identificar e quantificar os comentários odiosos na plataforma, investigamos também o conteúdo desses discursos para responder à QP3, por meio de modelagem de tópicos, técnica empregada para identificar padrões semânticos recorrentes e agrupar comentários com temáticas similares, permitindo interpretar quais formas de ofensa se manifestam com maior frequência no conjunto analisado. Para tanto, utilizamos o modelo *BERTopic* [16], método de modelagem de tópicos baseado em representações semânticas obtidas por modelos BERT. Em linhas gerais, o BERTopic transforma

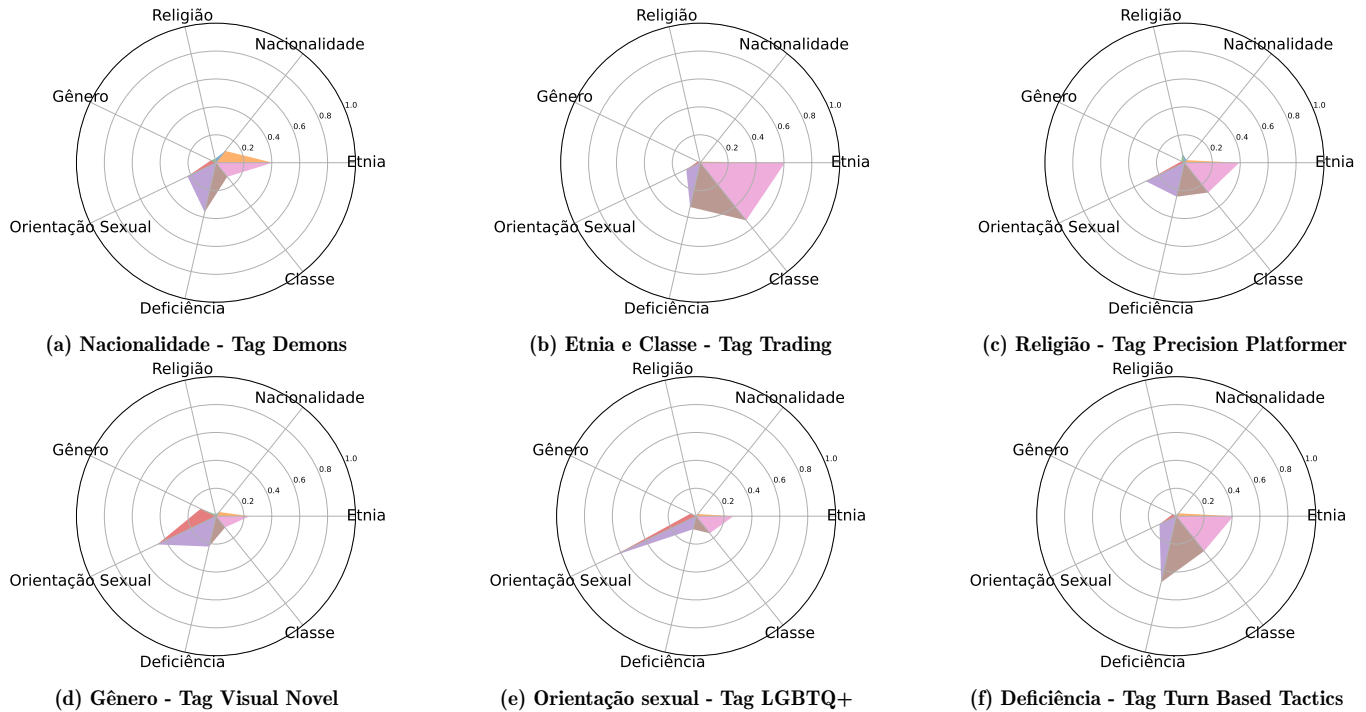


Figura 6: Anatomia do discurso para a Tags com maior prevalência de cada categoria de ódio.

os textos em vetores que capturam seu significado contextual, agrupa comentários semanticamente semelhantes por meio de algoritmos de clusterização e extrai palavras-chave representativas de cada tópico identificado.

A modelagem foi conduzida sobre os 13.811 comentários odiosos, dos quais o modelo atribuiu tópicos a 50,32%. A elevada variabilidade temática, decorrente da diversidade de jogos presentes no corpus, contribui para a dificuldade de atribuição dos demais comentários. A Tabela 3 apresenta os tópicos identificados, suas palavras-chave representativas e o volume de comentários associados a cada um. A nomenclatura foi gerada pelo modelo de linguagem Gemini a partir das palavras-chave extraídas pelo BERTopic.

A análise dos tópicos revela uma distribuição heterogênea do discurso de ódio, manifestando-se entre diferentes categorias temáticas. Nos tópicos de caráter mais abrangente, i.e., *Insultos Jogo*, *Problemas Técnicos* e *Violência Verbal*, observa-se o uso recorrente de termos como *trash*, *idiot* e *bird*, direcionados tanto à qualidade técnica dos jogos quanto aos seus desenvolvedores. Aprofundando a análise, evidencia-se a predominância do discurso de ódio relacionado à *etnia*, categoria mais recorrente entre os tópicos e presente também em comentários de jogos como *Smite* e *Garfield Kart*. Esse tipo de conteúdo concentra-se nos tópicos *Discurso Racista*, *Ódio Nacional* e *Insultos Raciais*, nos quais há clara intenção de hostilizar minorias étnicas, incluindo negros, asiáticos e eslavos. Termos como *russians* e *chinese* surgem frequentemente associados a expressões pejorativas como *trash*, enquanto termos como *negro* são mobilizados de forma ofensiva. Identifica-se ainda discurso de ódio direcionado a gênero nos

tópicos *Insultos Femininos* e *Degradação Sexual*, com manifestações voltadas predominantemente às mulheres, seja em jogos com protagonistas femininas, como *Emily is Away*, seja em contextos nos quais personagens femininas ocupam papéis secundários. Expressões como *slaves* e *sluts* ilustram a recorrência de linguagem que não apenas ofende, mas objetiva essas personagens e, por extensão, as mulheres.

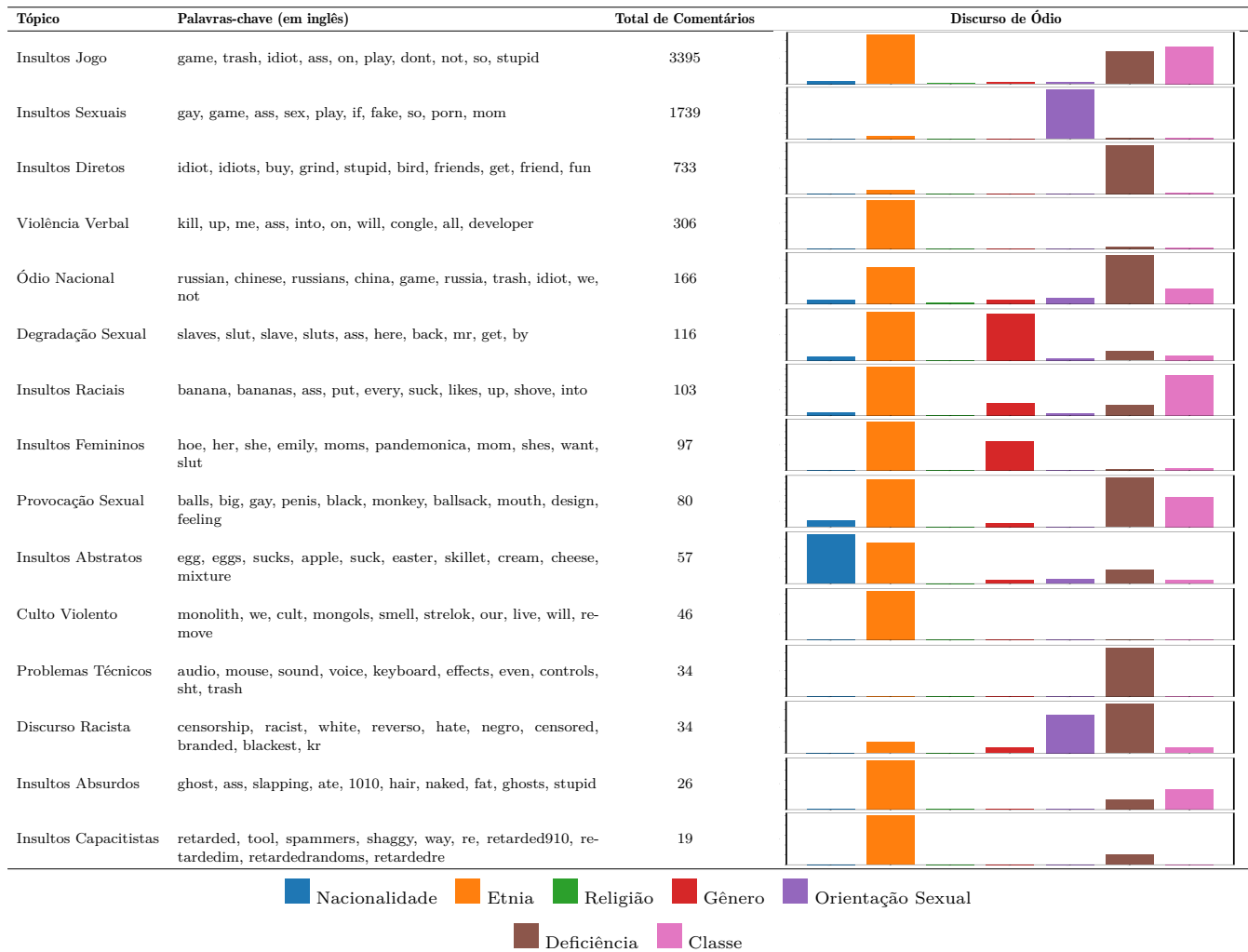
Observa-se também a incidência de discurso de ódio de natureza capacitista, em jogos como *Castle Crashers*, concentrado em tópicos como *Insultos Capacitistas*, *Culto Violento* e *Insultos Diretos*. Nesses agrupamentos, usuários recorrem a termos como *retarded*, *mongols* e *idiots* para desqualificar intelectualmente tanto outros jogadores quanto desenvolvedores. Observa-se igualmente a incidência de discurso capacitista, concentrada nos tópicos *Insultos Capacitistas*, *Culto Violento* e *Insultos Diretos*, com destaque para jogos como *Castle Crashers*. Nesses agrupamentos, termos como *retarded*, *mongols* e *idiots* são empregados para desqualificar intelectualmente tanto outros jogadores quanto desenvolvedores.

Por fim, o discurso de ódio relacionado à classe social, embora menos representativo em comparação às demais categorias, aparece de forma dispersa entre diferentes tópicos, frequentemente associado ao uso do termo *trash* em contextos que extrapolam a crítica e assumem conotação socioeconômica, sendo mobilizado para inferiorizar indivíduos com base em sua posição de classe.

7 CONCLUSÃO

Este trabalho se propõe a caracterizar o discurso de ódio na comunidade de jogos, sendo este estudo feito a partir da Steam, uma das maiores plataformas de distribuição de jogos

Tabela 3: BERTopic com distribuição das categorias de discurso de ódio.



do mundo. Devido ao seu alcance global e à variedade de espaços de socialização na plataforma, focamos em analisar e compreender o ódio presente nos comentários dos jogos adquiridos pelos usuários. Para isso, foi realizada uma vasta coleta dos comentários contidos na loja da plataforma. Com apoio de um léxico de termos de ódio (Hatebase) e de uma ferramenta para detecção de toxicidade (Perspective API) foi possível determinar onde, contra quem e de que forma aparecem os discursos odiosos, traçando assim uma anatomia desse tipo de discurso na loja da Steam.

Observando os resultados e as análises constata-se que apesar da moderação da plataforma ainda há uma presença demarcada nos comentários de discursos odiosos, estes muitas vezes mascarados por estratégias dos usuários de burlar a moderação. Buscando responder à **QP1** sobre onde é prevalente o discurso de ódio observamos uma série de jogos com uma proporção de ódio semelhante, contudo, destaca-se o jogo *Maize* apresentando uma proporção notoriamente maior de comentários odiosos e ao mesmo tempo concentrado em

um tipo de ódio: deficiência. Olhando para as *tags*, vemos que as mais tóxicas apresentam uma proporção mais próxima e com certa diversidade de discursos, sendo relevante ressaltar as *tags LGBTQ+* e *Trading* que figuram entre as mais odiosas e estão entre as que representam a maior proporção de comentários odiosos de determinados tipos.

Baseando-se nas categorias de ódio definidas pela Hatebase para responder à **QP2** sobre os alvos do ódio, observa-se que os tipos de ódio propagados concentram-se em etnia, orientação sexual, deficiência e classe, dos quais, etnia e classe concentram-se na *tag Trading*, orientação sexual na *LGBTQ+* e deficiência na *Turn Based Tactics*. Além disso, ao verificar os temas presentes nos discursos odiosos para responder à **QP3** notou-se uma presença marcante de 4 grandes grupos de tópicos: discurso de ódio geral, insultos capacitistas, discurso relacionado à orientação sexual e gênero, discurso racial e étnico.

LIMITAÇÕES E CUIDADOS ÉTICOS

Este trabalho apresenta limitações que devem ser consideradas na interpretação dos resultados. A análise restringiu-se ao idioma inglês, o que limita a generalização dos achados para outras culturas e idiomas presentes na plataforma. Outro ponto é o uso do léxico pré definido que pode não capturar todas as formas de expressar discursos odiosos, sobretudo aqueles com termos e técnicas para burlar moderação mais emergentes. Além disso, métodos automatizados de detecção de toxicidade podem apresentar dificuldades em capturar sarcasmos, ironias ou termos específicos de determinadas comunidades de jogos.

Em trabalhos futuros pretende-se explorar comentários em diferentes linguagens, sobretudo as que apresentam grandes volumes como chinês e português. Além disso, planeja-se estender a avaliação visando incluir novos léxicos de maneira a trazer resultados ainda mais completos e precisos. Somado a isto, pretende-se adicionar uma camada de validação humana no conjunto final de comentários a fim de aumentar a confiabilidade dos resultados.

No que tange aos cuidados éticos, o estudo atende à Resolução CNS nº 510/2016. A pesquisa utilizou exclusivamente dados públicos obtidos de forma legítima por meio da API oficial da Steam. Para garantir a proteção e em conformidade com a LGPD, todas as informações coletadas passaram por um processo rigoroso de anonimização, impedindo a identificação dos usuários nas análises realizadas, de forma agregada.

AGRADECIMENTOS

Os autores agradecem à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo suporte financeiro concedido para a realização desta pesquisa, por meio do Processo nº APQ-03115-24 Edital Demanda Universal. Além disso, Julio C. S. Reis agradece ao suporte do INCT-TILD-IAR (processo nº 408490/2024-1). Por fim, o uso de ferramentas de Inteligência Artificial generativa restringiu-se exclusivamente à revisão gramatical e ao aprimoramento da clareza do texto. A concepção do estudo, metodologia, condução dos experimentos, análise e conclusões são de responsabilidade integral dos autores.

REFERÊNCIAS

- [1] Shabana Arif. 2018. *Total War: Rome 2 Is Being Review Bombed on Steam over Imagined Controversy*. Acessado em: 9 feb. 2026.
- [2] Backlinko Team. 2026. Steam Usage and Catalog Stats. BackLinko. <https://backlinko.com/steam-users>. Online. Acessado em: 27 mar. 2026.
- [3] Beatriz Ribeiro Barreto et al. 2025. Indexação de mídias não convencionais: a recuperação de jogos eletrônicos na plataforma digital Steam. (2025).
- [4] Karin Becker. 2014. Introdução à Mineração de Opiniões. (2014).
- [5] Jeremy Blackburn and Haewoon Kwak. 2014. STFU NOOB! predicting crowdsourced decisions on toxic behavior in online games. In *Proceedings of the 23rd International Conference on World Wide Web (Seoul, Korea) (WWW '14)*. Association for Computing Machinery, New York, NY, USA, 877–888. doi:10.1145/2566486.2567987
- [6] Danah M. Boyd and Nicole B. Ellison. 2008. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication* 13 (2008), 210–230. doi:10.1111/j.1083-6101.2007.00393.x
- [7] David Buckingham. 2007. *Beyond Technology: Children's Learning in the Age of Digital Culture* (1st edition ed.). Polity, Cambridge.
- [8] Andy Chalk. 2018. *Total War: Rome 2 is getting review-bombed on Steam because of women generals (Updated)*. Retrieved Feb 9, 2026 from <https://www.pcgamer.com/total-war-rome-2-is-getting-review-bombed-on-steam-because-of-women-generals/>
- [9] Susana Costa, Bruno Mendes da Silva, and Mirian Tavares. 2023. Online hate speech in video games communities: A counter project. In *Counterspeech*. Routledge, 123–143.
- [10] Carolyn M. Cunningham. 2018. *Games Girls Play: Contexts of Girls and Video Games*. Lexington Books, Blue Ridge Summit.
- [11] Mai ElSherief, Shirin Nilizadeh, Dana Nguyen, Giovanni Vigna, and Elizabeth Belding. 2018. Peer to Peer Hate: Hate Speech Instigators and Their Targets. *Proceedings of the International AAAI Conference on Web and Social Media* 12 (04 2018). doi:10.1609/icwsm.v12i1.15038
- [12] Paula Fortuna and Sérgio Nunes. 2018. A Survey on Automatic Detection of Hate Speech in Text. *Comput. Surveys* 51, 4 (2018), 85:1–85:30. doi:10.1145/3232676
- [13] Jesse Fox and Wai Yen Tang. 2016. Women's experiences with general and sexual harassment in online video games. *New Media & Society* 19, 8 (2016), 1290–1307. doi:10.1177/1461444816635778
- [14] James Paul Gee. 2003. What video games have to teach us about learning and literacy. *Comput. Entertain.* 1, 1 (2003), 20. doi:10.1145/950566.950595
- [15] Paula Gíbrim, Marcus Ribeiro, Lorena Moreira, Philippe Melo, and Julio Reis. 2025. I'm not Welcome Here – Uma Revisão de Escopo sobre Toxicidade e Discurso de Ódio em Jogos Digitais. In *Proceedings of the 31st Brazilian Symposium on Multimedia and the Web (Rio de Janeiro/RJ)*. SBC, Porto Alegre, RS, Brasil, 646–660. doi:10.5753/webmedia.2025.15998
- [16] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794* (2022).
- [17] Samuel Guimaraes, Gabriel Kakizaki, Philippe Melo, Márcio Silva, Fabricio Murai, Julio CS Reis, and Fabrício Benevenuto. 2023. Anatomy of hate speech datasets: Composition analysis and cross-dataset classification. In *Proceedings of the 34th ACM Conference on Hypertext and Social Media*. 1–11.
- [18] Daniel G Heslep and PS Berge. 2024. Mapping Discord's darkside: Distributed hate networks on Disboard. *New Media & Society* 26, 1 (2024), 534–555. doi:10.1177/14614448211062548
- [19] György Kovács, Pedro Alonso, and Rajkumar Saini. 2021. Challenges of Hate Speech Detection in Social Media. *SN Computer Science* 2, 2 (Feb. 2021), 95. doi:10.1007/s42979-021-00457-3
- [20] Rachel Kowert, Emese Domahidi, Ruth Festl, and Thorsten Quandt. 2014. Social gaming, lonely life? The impact of digital game play on adolescents' social circles. *Computers in Human Behavior* 36 (2014), 385–390. doi:10.1016/j.chb.2014.04.003
- [21] Lucas Lima, Julio CS Reis, Philippe Melo, Fabricio Murai, Leandro Araujo, Pantelis Vikatos, and Fabricio Benevenuto. 2018. Inside the right-leaning echo chambers: Characterizing gab, an unmoderated social system. In *Proc. of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. 515–522.
- [22] Lucas Lima, Julio CS Reis, Philippe Melo, Fabricio Murai, and Fabrício Benevenuto. 2020. Characterizing (un) moderated textual data in social systems. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 430–434.
- [23] João Lobo, Daniel Barbosa, and Philippe Melo. 2025. Investigando a Dinâmica da Propagação do Ódio em Cascatas de Toxicidade nos Chats Ao Vivo da Twitch. In *Proceedings of the 31st Brazilian Symposium on Multimedia and the Web (Rio de Janeiro/RJ)*. SBC, Porto Alegre, RS, Brasil, 321–329. doi:10.5753/webmedia.2025.15110
- [24] Topias Mattinen and Joseph Macey. 2018. Online Abuse and Age in Dota 2. In *Proceedings of the 22nd International Academic Mindtrek Conference (Tampere, Finland) (Mindtrek '18)*. Association for Computing Machinery, New York, NY, USA, 69–78. doi:10.1145/3275116.3275149
- [25] Lorena Souza Moreira, Paula Teresa M. Gíbrim, Leonardo Rocha, and Julio C. S. Reis. 2026. Anatomy of Data Repositories for the Analysis and Detection of Toxicity in Portuguese. In *Proceedings of the 17th International Conference on Computational Processing of Portuguese - Vol. 1*. Association for Computational

- Linguistics, Salvador, BA.
- [26] Torill Elvira Mortensen. 2016. Anger, Fear and Games: The Long Event of #GamerGate. *Games and Culture* 13, 8 (2016), 787–806. doi:10.1177/1555412016640408
 - [27] Tathagata Mukherjee. 2023. Videogame distribution and Steam's imperialist practices: Platform coloniality in game distribution. *Journal of Games Criticism* 5 (2023).
 - [28] Newzoo. 2025. *Global Games Market Report*. Technical Report. Newzoo. <https://newzoo.com/resources/blog/global-games-market-report>
 - [29] Rogerio Nunes and Ivelise Fortim. 2025. Oportunidades, riscos e danos nas plataformas adjacentes a games. In *Anais do XXIV Simpósio Brasileiro de Jogos e Entretenimento Digital* (Salvador/BA). SBC, Porto Alegre, RS, Brasil, 792–803. doi:10.5753/sbgames.2025.9558
 - [30] Samuel Lopes Pinto, José Julio Campolina, Joao Pedro M Sena, Gabriel Félix, Lucas N Ferreira, and Julio CS Reis. 2024. Caracterização e predição de usuários tóxicos no twitter/x durante as eleições brasileiras de 2022. In *Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)*. 61–74.
 - [31] Luke Plunkett. 2013. *Steam Is 10 Today. Remember When It Sucked?* Acessado em: 4 fev. 2026.
 - [32] MARC PRENSKY. 2006. *Don't Bother Me Mom - I'm Learning!: How Computer And Video Games Are Preparing Your Kids for Twenty-First Century Success - And How You Can Help!* (illustrated edição ed.). PARAGON HOUSE, St. Paul, Minn.
 - [33] Robert D. Putnam. 2000. *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster, New York.
 - [34] Marcus Vinicius Guerra Ribeiro, Clara Andrade Pimentel, and Philipe de Freitas Melo. 2024. Quando as Avaliações Viram Bombas: Explorando a Dinâmica do Review Bombing nos Jogos no Metacritic. In *Anais do XXX Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia)*. Sociedade Brasileira de Computação (SBC), Juiz de Fora, MG, Brasil, 249–256. doi:10.5753/webmedia.2024.241744
 - [35] u/diogenesl. 2018. *Total War: Rome 2 is getting review-bombed on Steam because of women generals*. Acessado em: 9 fev. 2026.
 - [36] Job Valkenburg. 2026. Discover the steam community: enhance your gaming experience. <https://unityagency.com/en/steam-community/>. Acessado em: 14 mar. 2026.
 - [37] xPaw. 2026. Steam Charts. <https://steamdb.info/app/753/charts/>. Acesso em: 04 fev. 2026.