

I'm not Welcome Here – Uma Revisão de Escopo sobre Toxicidade e Discurso de Ódio em Jogos Digitais

Paula Teresa M. Gibrim

paula.gibrim@ufv.br

Universidade Federal de Viçosa
Viçosa, Minas Gerais, Brasil

Marcus Vinicius Guerra Ribeiro

marcus.guerra@ufv.br

Universidade Federal de Viçosa
Viçosa, Minas Gerais, Brasil

Lorena Souza Moreira

lorenasouzamoreiraa@gmail.com

Universidade Federal de Viçosa
Viçosa, Minas Gerais, Brasil

Philippe de F. Melo

philipe.freitas@ufv.br

Universidade Federal de Viçosa - CAF
Florestal, Minas Gerais, Brasil

Julio C. S. Reis

jreis@ufv.br

Universidade Federal de Viçosa (UFV)
Viçosa, Minas Gerais, Brasil

ABSTRACT

Hate speech and toxic behaviors are increasingly common in digital games, affecting player wellbeing, diversity and community engagement. Although research on this topic has increased in recent years, studies remain fragmented and largely focused on the North American and European contexts. This study presents a scoping review of the scientific literature on hate speech in games, covering publications from 2010 to 2025, including international and Brazilian research. A search of four electronic library sources, complemented by manual searches, retrieved 8,040 records. After filtering and screening using predefined criteria, 49 studies were selected for full-text analysis. The review presents the investigated associations, detection approaches and intervention strategies, consolidates the scattered results and identifies areas for further investigation. Among these observations is the relative scarcity of research focused on Brazilian gaming communities, suggesting opportunities for more regionally contextualized studies in the future.

KEYWORDS

games, jogos, video games, videogames, *digital games*, jogos digitais, *hate speech*, discurso de ódio, *toxicity*, toxicidade, *scope review*, revisão de escopo, *literature review*, revisão da literatura

1 INTRODUÇÃO

Os jogos digitais têm se consolidado como um dos pilares centrais da indústria do entretenimento global [87], superando até mesmo a do cinema e ocupando um papel que vai além do próprio jogo: eles funcionam como espaços de interação social, formação de comunidades e expressão cultural, conectando bilhões de pessoas em diferentes plataformas e contextos [23, 80]. Conforme o relatório *Global Games Market Report 2024* [60], da Newzoo, o mercado de jogos gerou uma receita de US\$ 187,7 bilhões em 2024, com uma base global de 3,42 bilhões de jogadores, demonstrando a relevância crescente dos jogos digitais como fenômeno econômico e cultural.

Esse crescimento expressivo não se limita ao consumo de jogos, mas reflete uma participação cultural mais ampla. A cultura

dos jogos digitais transcende o *software*, manifestando-se em ecossistemas de redes sociais, fóruns e plataformas de *streaming* que funcionam hoje como extensões do próprio ambiente de jogo [34]. Nesses espaços, jogadores compartilham experiências, discutem estratégias, celebram conquistas e, em muitos casos, prolongam dinâmicas competitivas e sociais iniciadas durante as partidas [27]. Essas práticas sustentam comunidades engajadas, nas quais são construídos significados coletivos, formas de pertencimento e normas de conduta compartilhadas. Contudo, essas interações não ocorrem sem seus problemas: elas são frequentemente atravessadas por manifestações de toxicidade [36, 80, 82].

No contexto dos jogos digitais, o termo “toxicidade” é frequentemente empregado, tanto na literatura acadêmica quanto na mídia, como um conceito guarda-chuva que abrange diferentes práticas nocivas. Essas práticas vão desde condutas que comprometem a experiência coletiva, como comportamentos desrespeitosos, disruptivos ou de sabotagem intencional, até manifestações discriminatórias, como assédio, racismo e machismo [9, 19, 86]. Dentro desse espectro mais amplo, o discurso de ódio se configura como uma das formas mais graves de toxicidade, por envolver ataques direcionados a indivíduos ou grupos com base em características identitárias. Como destaca a Organização das Nações Unidas para a Educação, Ciência e Cultura (UNESCO), tais comportamentos reforçam estereótipos e desigualdades, restringem a participação de grupos vulneráveis e enfraquecem o caráter democrático dos espaços digitais [33].

Embora amplamente discutido, o discurso de ódio ainda carece de uma definição universalmente aceita [32] e levanta bastante debate na sociedade. Neste estudo, adota-se a definição proposta por Fortuna e Nunes [2018], segundo a qual se trata de qualquer forma de linguagem que ataca, diminui ou incita ódio ou violência contra indivíduos ou grupos, com base em características como aparência física, religião, origem étnica ou nacional, orientação sexual ou identidade de gênero. Assim, no contexto dos jogos digitais, consideramos como discurso de ódio, e, portanto, como uma manifestação específica de toxicidade, qualquer comunicação escrita, verbal ou comportamental que promova esse tipo de ataque ou discriminação. Essa definição inclui tanto manifestações explícitas quanto sutis, inclusive aquelas disfarçadas sob a forma de humor, reconhecendo a complexidade das interações online e a importância de compreender como tais práticas afetam a convivência e a participação de diferentes grupos nos espaços relacionados aos jogos.

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'2025). Rio de Janeiro, Brazil. Porto Alegre: Brazilian Computer Society, 2025.
© 2025 SBC – Brazilian Computing Society.
ISSN 2966-2753

Um exemplo emblemático dessa manifestação odiosa foi evidenciada pela campanha #MyGameMyName [13], que destacou a dimensão do assédio enfrentado por jogadoras. A iniciativa partiu de um dado alarmante da Newzoo [59]: embora as mulheres representem 46% do público *gamer*, muitas optam por ocultar seu gênero para evitar ataques. Para expor essa realidade, a Agência Africa¹ e a organização Wonder Woman Tech² convidaram influenciadores homens a jogar utilizando nomes de usuário e avatares femininos. Ao vivenciarem, em primeira mão, o volume de insultos e ameaças dirigidos às mulheres, esses criadores de conteúdo compartilharam suas experiências com milhões de seguidores, ampliando o debate sobre o problema. A campanha ganhou reconhecimento internacional ao vencer o prêmio Glass do Cannes Lions [74], dedicado a iniciativas que promovem a equidade de gênero. Apesar da repercussão, o caso não representa uma vitória, mas sim um alerta para a gravidade e crescimento contínuo desse fenômeno. A propagação de ódio em jogos segue como um desafio persistente, exigindo ações coordenadas de comunidades, pesquisadores e desenvolvedores para mitigar seus impactos e promover ambientes virtuais mais inclusivos e seguros.

Apesar do crescimento das discussões sobre toxicidade e discurso de ódio em jogos digitais, a produção científica sobre o tema permanece fortemente concentrada em países do hemisfério norte, com publicações em inglês e análises voltadas a comunidades próprias. Em contraste, o Brasil, mesmo sendo um dos dez maiores mercados de jogos do mundo [60], ainda carece de uma centralização dos estudos que investiguem como essas dinâmicas se manifestam em seu contexto sociocultural. Essa lacuna representa um déficit de representatividade acadêmica e compromete diretamente a construção de ambientes virtuais mais inclusivos e seguros, já que a ausência de dados e estratégias adaptadas à realidade brasileira deixa comunidades locais vulneráveis à reprodução e, em muitos casos, à amplificação das mesmas estruturas de preconceito e desigualdade que permeiam a sociedade *offline*. Essa carência de investigações se torna ainda mais preocupante diante da dimensão da comunidade *gamer* no país. Segundo a Pesquisa Game Brasil (PGB) de 2024 [79], 73% dos brasileiros jogam com regularidade, destacando a importância de espaços culturais para os jogos digitais no Brasil. Essa expressiva participação evidencia a urgência de compreender melhor como o discurso de ódio se manifesta e é tratado nessas comunidades, de modo que resultados obtidos em estudos internacionais possam ser devidamente contextualizados e, quando necessário, adaptados à realidade sociocultural brasileira.

Dante do exposto, o objetivo geral deste trabalho é mapear e analisar, como o discurso de ódio no contexto dos jogos digitais tem sido estudado e abordado na literatura científica. Para isso, conduziu-se uma revisão de escopo que considera pesquisas nacionais e internacionais. Primeiramente, buscou-se levantar e classificar os estudos existentes quanto aos tipos de jogos e plataformas mais investigados. Em seguida, foram identificados os métodos empregados para detectar e analisar a toxicidade. Por fim, mapeou-se as estratégias de mitigação mais recorrentes propostas na literatura, englobando soluções técnicas, de *design* e políticas comunitárias. Cabe destacar também que esta revisão apresenta um enfoque predominantemente

quantitativo, priorizando a sistematização e análise de tendências mensuráveis nos estudos revisados.

Em resumo, a sistematização desses resultados, além de permitir uma compreensão do estado atual da pesquisa sobre o tema, também evidencia lacunas que podem orientar investigações futuras e auxiliar na formulação de estratégias para comunidades e desenvolvedores interessados em criar espaços mais inclusivos e seguros.

O restante deste trabalho está estruturado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados ao tema, que fornecem a base conceitual e contextual para a pesquisa sobre discurso de ódio em jogos; a Seção 3 descreve detalhes relativos à metodologia empregada na condução da revisão de escopo; depois, a Seção 4 apresenta as análises e discussões dos dados obtidos ao longo do estudo; por fim, a Seção 5 reúne as conclusões e reflexões sobre o panorama identificado.

2 TRABALHOS RELACIONADOS

A pesquisa sobre o discurso de ódio em ambientes digitais, especialmente no contexto dos jogos eletrônicos, vem ganhando relevância nas últimas décadas. Um conceito central para compreender a naturalização de comportamentos hostis nesses espaços é o de violência simbólica, formulado por Pierre Bourdieu [85]. Trata-se de uma forma sutil, muitas vezes invisível, de dominação que se manifesta por meio da linguagem, de representações culturais e de normas sociais que reforçam desigualdades e exclusões. No contexto dos jogos, essa violência pode ser observada tanto nas dinâmicas de interação entre jogadores quanto nas representações de gênero, raça e classe veiculadas em jogos e plataformas de *streamings*.

Um aspecto amplamente discutido na literatura sobre toxicidade em jogos é a normalização por parte dos usuários, e consequente dessensibilização, frente a comportamentos discriminatórios e agressivos nesses ambientes. Como apontam Beres *et al.* [2021], práticas misóginas e agressões verbais deixam de ser percebidas como transgressões e passam a ser incorporadas à chamada “cultura *gamer*”, muitas vezes sob o pretexto de autenticidade competitiva. Nesse cenário, muitos jogadores acabam aceitando insultos e hostilidades como “parte do jogo”, reproduzindo e tolerando dinâmicas que dificilmente seriam aceitáveis em interações fora do ambiente virtual. Essa naturalização também se manifesta em experiências específicas de minorias. Ortiz [2019] investiga como homens não brancos enfrentam episódios constantes de racismo em jogos online, revelando que a dessensibilização surge não como indiferença, mas como uma estratégia de sobrevivência emocional em um espaço que normaliza a propagação do ódio. Essa resposta implica custos psicológicos significativos e perpetua um ciclo em que a hostilidade permanece invisível ou pouco contestada.

Essa lógica de adaptação também é observada em contextos competitivos. No estudo de Türkay *et al.* [2020], jogadores universitários de *e-sports*³ relatam que a toxicidade é definida menos pelo conteúdo ofensivo e mais pelo impacto na moral e na coesão da equipe. Para continuar competindo, eles adotam estratégias de enfrentamento passivas, como ignorar os ofensores (*See No Evil*), silenciar a comunicação (*Hear No Evil*) e evitar confrontos (*Speak No Evil*).

¹Acessível em: <https://africacreative.com.br/>

²Acessível em: <https://wonderwomentech.com>

³O termo *e-sports* define a modalidade competitiva de jogos virtuais, onde jogadores profissionais ou amadores se enfrentam em torneios oficiais [10].

Embora essas práticas funcionem como mecanismos de autopreservação, acabam por reforçar a naturalização da perpetuação do ódio como parte integrante da cultura competitiva, dificultando esforços de mitigação mais estruturais.

Essa exposição contínua a comportamentos hostis também afeta diretamente a permanência e o engajamento dos jogadores. Estudos como os de Kwak *et al.* [2015] e Reid *et al.* [2022b] apontam que ambientes marcados pela toxicidade reduzem o senso de bem-estar e controle dos participantes, gerando frustração e evasão, especialmente entre jogadores novatos ou menos tolerantes a conflitos. Kwak *et al.* [2015] destacam que a intensidade da linguagem abusiva se correlaciona com o abandono de jogos, evidenciando que, em muitos casos, a hostilidade prejudica a experiência individual e compromete a sustentabilidade das comunidades virtuais.

Essa lógica de hostilidade e evasão não se limita às interações diretas entre jogadores, mas também se manifesta em práticas coletivas de boicote e retaliação que vão para além do jogo, como o *review bombing*, que é a prática em que uma comunidade se mobiliza para inundar um produto (como jogo, filme, livro, entre outros) com avaliações negativas ou positivas, visando alterar a sua percepção pública [62, 67]. Ribeiro *et al.* [2024] analisam como essa dinâmica, amplificada em plataformas como o Metacritic⁴, opera como uma forma de pressão social e simbólica que pode reforçar tensões dentro das comunidades de jogadores. Embora o *review bombing* não seja, por si só, uma prática sempre ligada ao discurso de ódio, sua associação com pautas de intolerância e reações extremas a questões de diversidade em jogos evidencia como manifestações hostis podem extrapolar o ambiente de jogo e atingir outros espaços de interação e influência cultural.

Em uma escala ainda mais ampla, Chatzakou *et al.* [2017] discutem o fenômeno conhecido como *GamerGate* [18], amplamente considerado um marco na hostilidade dirigida a mulheres no ecossistema dos jogos. Emergido em 2014, o movimento rapidamente se revelou uma campanha de assédio sistemático direcionada a mulheres desenvolvedoras, jornalistas e outras figuras públicas do setor. Mais do que um episódio isolado, o *GamerGate* evidenciou o potencial dos ecossistemas *gamer* para organizar e coordenar ataques misóginos, sendo frequentemente apontado como um ponto de inflexão na discussão global sobre cultura dos jogos, assédio e discurso de ódio online. Essa conjuntura é aprofundada pela obra *Games Girls Play*, de Cunningham [2018], que investiga como meninas e mulheres vivenciam os jogos e suas comunidades. A autora evidencia que essas jogadoras enfrentam tanto barreiras explícitas (como assédio e insultos diretos) quanto implícitas (como exclusão velada e sub-representação em narrativas e papéis de destaque). Muitas acabam recorrendo a estratégias de autoproteção ou abandonando comunidades hostis.

Enquanto muitos estudos que se concentram majoritariamente no conteúdo tóxico produzido pelos jogadores ou nos usuários envolvidos tanto como autores ou vítimas dos comportamentos abusivos, Pimentel e Melo [2020] direcionam o foco para o próprio sistema de jogo, analisando como elementos de *game design* podem favorecer a toxicidade. Com base em fatores de desinibição tóxica, os autores identificaram como mecânicas que estimulam a excessiva individualização dos jogadores, o enfraquecimento de

laços de solidariedade e a indiferença em relação a comportamentos nocivos contribuem para o agravamento de ofensas e conflitos, e como ferramentas disponibilizadas pelos jogos são muitas vezes ineficazes no combate da toxicidade. Esse enfoque amplia o debate ao responsabilizar também as escolhas de arquitetura e *design* dos jogos na formação de dinâmicas comunitárias que podem estimular ou mitigar comportamentos tóxicos, conectando-se a uma linha de pesquisa que propõe olhar para o papel das plataformas e mecânicas de jogo como agentes ativos na construção de interações.

A persistência dessas tensões sociais evidencia a necessidade de compreender e também intervir de forma efetiva na cultura de toxicidade dos jogos digitais. Nesse sentido, a revisão sistemática de Wijkstra *et al.* [2024] oferece uma contribuição fundamental ao mapear e analisar as principais propostas de mitigação da toxicidade em jogos online. O estudo identifica uma predominância de abordagens punitivas na literatura, como sistemas de denúncia e banimentos automáticos, mas argumenta que tais mecanismos, embora necessários, não são suficientes para promover mudanças culturais duradouras. Os autores destacam a urgência de estratégias mais holísticas e contextuais que envolvam, por exemplo, intervenções de *design* e programas de educação comunitária e apontam uma lacuna crítica: a maioria das soluções propostas academicamente não é testada ou aplicada em cenários reais, limitando sua eficácia prática.

Apesar dos avanços nos estudos sobre discurso de ódio em jogos digitais, a literatura ainda carece de uma visão integrada. As discussões permanecem fragmentadas e sem uma articulação clara sobre como plataformas, comunidades e interações moldam o fenômeno. Poucos trabalhos analisam como essas comunidades se estruturam, entretanto, compreender essas dinâmicas é essencial para revelar como poder e vieses circulam e favorecem a propagação de comportamentos tóxicos e discursos hostis.

Este trabalho busca reduzir essa fragmentação ao mapear a produção científica nacional e internacional, com atenção especial ao contexto brasileiro — frequentemente negligenciado em revisões internacionais por barreiras linguísticas, limitações de indexação e dispersão em trabalhos isolados — por meio de uma revisão de escopo. Além disso, este trabalho busca não apenas sistematizar onde e como o discurso de ódio em jogos digitais tem sido estudado, mas também exaltar a relevância cultural e social desse campo, identificando as abordagens utilizadas para detectá-lo e as estratégias propostas para mitigá-lo, de modo a oferecer uma visão consolidada e acessível que contribua para o avanço da área.

3 METODOLOGIA

Essa pesquisa consiste em uma revisão de escopo, conduzida com base nas diretrizes metodológicas propostas por Arksey e O'Malley [2005]. Diferentemente de uma revisão sistemática da literatura convencionalmente aplicada, essa abordagem possui caráter mais exploratório e abrangente, sendo adequada para mapear e descrever a literatura existente sobre um tema, identificar lacunas e tendências e responder a questões de pesquisa menos específicas. Além disso, seguiram-se as orientações as orientações de relato estabelecidas pelo *Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews* (PRISMA-ScR) [81]. O processo metodológico adotado, ilustrado de forma resumida na Figura 1, contemplou as seguintes etapas: i) formulação das questões

⁴Acessível em: <https://www.metacritic.com/>

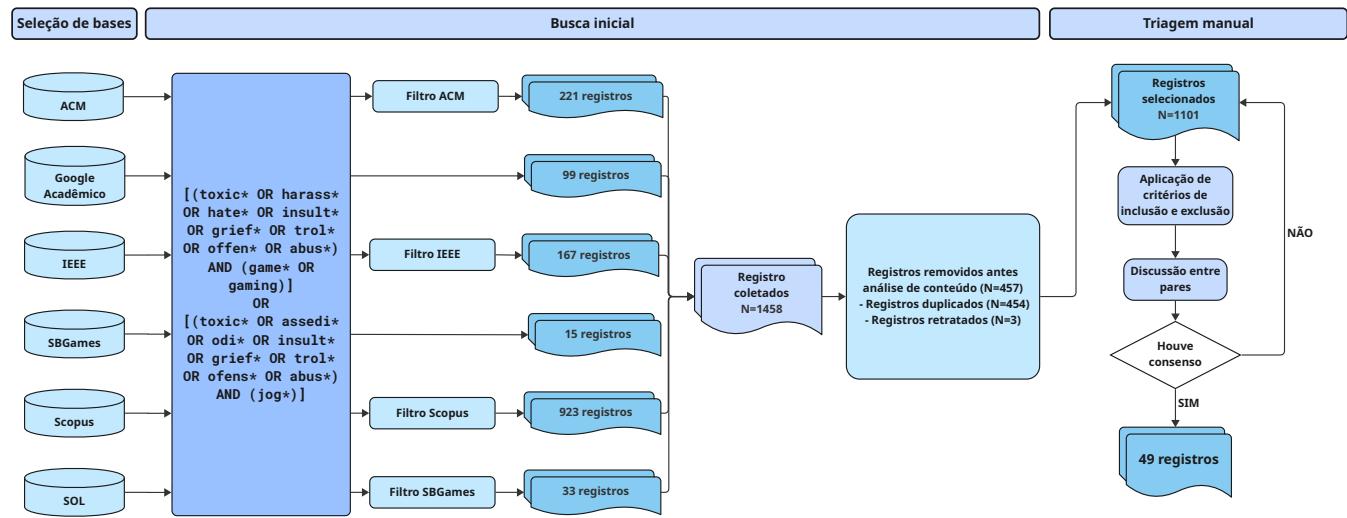


Figura 1: Fluxograma do processo de busca, filtragem e seleção dos estudos incluídos na revisão de escopo apresentada neste estudo, adaptado ao formato PRISMA-ScR.

de pesquisa (Seção 3.1); ii) seleção dos repositórios utilizados para a busca bibliográfica (Seção 3.2); iii) definição dos termos de busca (Seção 3.3); iv) definição dos critérios de inclusão (Seção 3.4); v) execução das buscas e recuperação dos registros (Seção 3.5); vi) triagem dos documentos segundo os critérios estabelecidos (Seção 3.6); e, vii) análise dos vieses e limitações metodológicas (Seção 3.7).

3.1 Questões de pesquisa

As questões de pesquisa foram elaboradas para mapear como a literatura científica tem abordado o fenômeno do discurso de ódio em jogos digitais, de modo a evidenciar tendências, lacunas e oportunidades para futuras investigações. Especificamente, esta revisão busca responder às seguintes questões (QPs):

- QP1:** Quais ambientes e contextos têm sido foco das análises de discurso de ódio em jogos digitais?
- QP2:** Quais técnicas e abordagens têm sido utilizadas para a detecção do discurso de ódio nesses ambientes de jogos?
- QP3:** Quais propostas de intervenção têm sido discutidas voltadas à mitigação do discurso de ódio em jogos?

3.2 Seleção das bases para busca

Para garantir abrangência e representatividade na identificação dos estudos, foram selecionadas quatro bibliotecas eletrônicas com escopos complementares: *IEEE Xplore*, *The ACM Full-Text Collection*, *Scopus* e a *SBC-OpenLib*. As bases *IEEE Xplore* e *The ACM Full-Text Collection* foram priorizadas por concentrarem a produção científica de maior relevância nas áreas de computação, jogos digitais e interação humano-computador. Nelas, são indexadas conferências e periódicos onde são publicados, com maior frequência e impacto, estudos sobre técnicas de detecção, análise e mitigação de toxicidade e discurso de ódio. De forma complementar, a *Scopus* foi incorporada

por sua ampla abrangência e caráter multidisciplinar, possibilitando a inclusão de estudos potencialmente relevantes que poderiam ser excluídos caso apenas bases especializadas fossem consideradas.

A *SBC-OpenLib* (Biblioteca Digital da Sociedade Brasileira de Computação) foi incorporada com o objetivo garantir representatividade da produção nacional, uma vez que é o principal repositório de conferências brasileiras de Computação. Contudo, em contraste com as bases internacionais, a recuperação de trabalhos brasileiros apresenta desafios específicos: sua indexação permanece incompleta e os mecanismos de filtragem são limitados, de tal forma a comprometer a precisão das buscas e fragmentar a área no país. Diante disso, se fez necessário a adoção de estratégias complementares para assegurar a inclusão de estudos relevantes nacionalmente, como a busca manual nos anais do SBGames – Simpósio Brasileiro de Jogos e Entretenimento Digital⁵ –, principal evento nacional da área. Essa busca abrangeu as edições anteriores a 2021, uma vez que essas edições não estão disponíveis em nenhuma base eletrônica. Cabe destacar que parte desse acervo permanece inacessível, como no caso dos anais do SBGames 2013, atualmente indisponíveis online.

Por fim, para ampliar a cobertura e reduzir o risco de omissão de estudos relevantes que pudesse ter sido indexado em outras bases, foi conduzida uma busca complementar no Google Acadêmico⁶, limitada às cinco primeiras páginas de resultados e utilizando os termos definidos para a busca bibliográfica. Essa abordagem combinada visou assegurar maior abrangência e completude na recuperação dos estudos.

3.3 Definição dos termos de busca

Conforme discutido anteriormente, a comunidade de jogadores ultrapassa os limites dos próprios jogos e, por essa razão, os termos

⁵Acessível em: <https://www.sbgames.org/>

⁶Acessível em: <https://scholar.google.com/>

Tabela 1: Critérios de inclusão e exclusão adotados na triagem dos registros identificados na revisão de escopo.

#	Inclusão	#	Exclusão
I.1	Publicações em português ou inglês;	E.1	Registros duplicados;
I.2	Artigos completos classificados como periódicos ou conferências;	E.2	Registros retratados;
I.3	Publicados entre janeiro de 2010 e maio de 2025;	E.3	Foco em áreas não relacionadas, como Educação ou Saúde;
I.4	Publicações na área da Computação;	E.4	Versões preliminares de trabalhos completos;
I.5	Estudo discute manifestações de discurso de ódio em jogos digitais ou em comunidades gamer;	E.5	Trabalhos em andamento (sem dados ou análise concluída);
I.6	Estudos que contribuem para responder ao menos uma das perguntas de pesquisa;	E.6	Estudo não contribui para responder às perguntas de pesquisa;
		E.7	Texto completo indisponível;

de busca foram definidos de modo a abranger expressões diretamente relacionadas ao discurso de ódio, além de comportamentos disruptivos que, ao serem contextualizados, funcionam como gatilhos ou desdobramentos dessas manifestações, mesmo em espaços externos aos jogos, mas relacionados ao seu ecossistema. Tendo isso em vista, e em acordo com os termos utilizados por Wijkstra *et al.* [2024], foram incluídas palavras-chave relacionadas ao discurso de ódio e à comportamentos tóxicos de maneira geral: “toxicidade”, “assédio”, “ódio”, “insulto”, “griefing”⁷, “troll”⁸, “ofensas” e “abuso”. Em relação ao contexto dos jogos, optou-se por não incluir termos que limitassem excessivamente o escopo, como “online”, “competitivo”, “cooperativo”, “e-sport”, “MOBA” (*Multiplayer Online Battle Arena*), “MMOG” (*Massive Multiplayer Online Game*) ou “FPS” (*First-person Shooter*), de modo a garantir uma revisão abrangente, independente do tipo ou modalidade de jogo analisada.

Para maximizar o alcance das buscas e contemplar variações morfológicas e flexões verbais, foram empregados caracteres curinga (*wildcards*). Essa estratégia permite que um único termo de busca abranja diferentes derivações lexicais. Por exemplo, ao utilizar “insult*”, o mecanismo de busca pode recuperar variações como “insultado”, “insultar” e “insulto”, entre outras formas relacionadas. Tal estratégia foi construída em duas versões – uma em inglês e outra em português – e ambas foram aplicadas em todas as bases consultadas, limitando-se aos campos de título (*title*), resumo (*abstract*) e palavras-chave (*keywords*). As strings finais foram:

Inglês: (toxic* OR harass* OR hate* OR insult* OR grief* OR troll* OR offend* OR abus*) AND (game* OR gaming)

Português: (toxic* OR assedi* OR odi* OR insult* OR grief* OR troll* OR ofens* OR abus*) AND (jog*)

3.4 Definição dos critérios de inclusão e exclusão

A seleção dos estudos foi realizada por três revisores independentes, com base em critérios de inclusão e exclusão previamente definidos

⁷O termo *griefing* refere-se a comportamentos intencionais adotados por jogadores com o objetivo de atrapalhar ou irritar outros participantes em um jogo, frequentemente sem buscar vantagens competitivas [24].

⁸O termo *troll* é usado para descrever indivíduos que provocam, insultam ou causam transtornos intencionais em interações online, seja para entreter a si mesmos, desestabilizar outros jogadores ou gerar conflitos [22].

conforme apresentado na Tabela 1. Esses critérios foram elaborados de modo a garantir que os estudos analisados estivessem alinhados aos objetivos da revisão e às questões de pesquisa, estabelecendo um recorte que prioriza publicações da área de Computação capazes de fornecer informações relevantes sobre a compreensão, detecção e mitigação do discurso de ódio em jogos digitais.

3.5 Execução da busca

As duas versões dos termos de busca foram adaptadas conforme a sintaxe de cada base e utilizadas para recuperar documentos cuja ocorrência dos termos estivesse nos campos de título (*title*), resumo (*abstract*) ou palavras-chave (*keywords*).

As buscas foram realizadas durante a segunda quinzena do mês de maio de 2025, sendo a última atualização conduzida em 05 de junho de 2025 e recuperou um total inicial de 8.040 registros. Todos os registros foram importados para o *Zotero Reference Manager*⁹, um gerenciador de referências gratuito e de código aberto que organiza, armazena e formata citações e bibliografias, que identificou automaticamente 8 itens retratados e 481 registros duplicados. Restaram, portanto, 7.551 documentos únicos.

Para otimizar a análise e reduzir o volume de registros a serem avaliados manualmente, foram utilizados, quando disponíveis, filtros automáticos oferecidos pelas próprias bases, sempre em alinhamento com os critérios de inclusão previamente definidos. As buscas conduzidas de forma manual, como no caso dos anais do SBGames e do Google Acadêmico, não passaram por essa filtragem automatizada. Ao final desse processo, que está descrito a seguir, foram selecionados 1.458 registros, dos quais 454 eram duplicados e 3 foram retratados, sobrando, assim, 1.101 registros únicos¹⁰ para análise manual.

3.5.1 Scopus. Por ser a base responsável pelo maior volume de resultados recuperados, foram aplicados múltiplos filtros: intervalo temporal (2010 a 2025, conforme I.3), área do conhecimento (Ciência da Computação, conforme I.4), tipo, estágio e fonte dos documentos (apenas artigos e trabalhos completos publicados em periódicos ou anais de conferência, conforme I.2), e idioma (inglês e português, conforme I.1). Essa etapa reduziu o total de 7.422 registros iniciais para 923 documentos a serem analisados manualmente, excluindo automaticamente 6.499 registros fora do escopo.

⁹Disponível em: <https://www.zotero.org/>

¹⁰Esses registros estão listados em: https://www.zotero.org/groups/6064544/hate_speech_in_digital_games_-_w_filters.

3.5.2 IEEE Xplore. Na *IEEE Xplore*, os filtros disponíveis eram mais limitados em relação aos critérios de inclusão definidos. Dessa forma, foi definido o intervalo temporal (2010 a 2025, conforme I.3) e a seleção exclusiva de registros classificados como *Conferences* e *Journals* (conforme I.2). Com isso, dos 204 registros inicialmente recuperados, 38 foram excluídos automaticamente, resultando em 167 documentos para análise manual.

3.5.3 The ACM Full-Text Collection. Por se tratar de uma base composta apenas por textos completos, a *ACM Full-Text Collection* já atendia ao critério relacionado à completude dos trabalhos (I.2). Dessa forma, foi aplicado apenas o filtro de intervalo temporal (2010 a 2025, conforme I.3). Dos 247 registros inicialmente recuperados, 26 foram eliminados automaticamente, restando 221 documentos para a etapa de análise manual.

3.5.4 SBC-OpenLib. Na *SBC-OpenLib*, a aplicação de filtros teve impacto irrelevante sobre o conjunto de resultados. Embora tenha sido utilizado o filtro de intervalo temporal (2010 a 2025, conforme I.3), todos os 33 registros retornados inicialmente pela busca já atendiam a esse critério, não havendo redução no volume inicial.

3.6 Triagem dos estudos

A partir desse conjunto, foi conduzida a triagem manual em múltiplas fases. Na primeira delas, procedeu-se à análise dos títulos, com foco na identificação de registros que atendessem aos critérios de idioma (I.1) e de alinhamento temático com o escopo da revisão (I.5, E.3). Essa filtragem resultou na exclusão criteriosa de 927 registros, reduzindo o conjunto para 174 estudos. Em seguida, foi realizada uma avaliação detalhada dos metadados – como a data e fonte da publicação –, e dos resumos, complementada, quando necessário, pela leitura das seções introdutórias dos textos, de modo a esclarecer dúvidas sobre a relevância dos estudos e garantir a aplicação integral dos critérios estabelecidos (Tabela 1). Ao final desse processo, foram selecionados 49 trabalhos que cumpriam todos os critérios de inclusão e exclusão, os quais seguiram para a leitura completa. Após o fim da triagem, os resultados foram discutidos entre os revisores para validação das decisões e resolução de divergências, assegurando a consistência do processo antes do avanço para a etapa de análise.

3.7 Avaliação de viés

Embora existam limitações metodológicas inerentes às decisões ao longo desta revisão, cada uma delas foi tomada com base em objetivos estratégicos. A limitação a estudos publicados em português e inglês, por exemplo, pode ter levado à exclusão de trabalhos relevantes publicados em outros idiomas. No entanto, essa escolha buscou contemplar a literatura internacional predominante, ao mesmo tempo em que ampliava a visibilidade de pesquisas desenvolvidas no contexto brasileiro, principalmente considerando a relevância do tema no cenário nacional.

De maneira semelhante, a opção por considerar apenas artigos completos publicados em periódicos ou anais de conferências levou à exclusão de teses, dissertações e outros tipos de publicações que poderiam trazer contribuições relevantes. No entanto, grande parte dessas produções acadêmicas, especialmente as de maior impacto, costuma ser convertida em artigos científicos, o que permite que

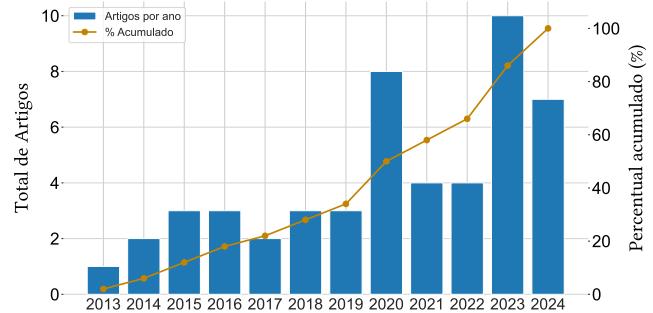


Figura 2: Distribuição temporal dos estudos incluídos na revisão, indicando o número de publicações por ano e o percentual acumulado no período analisado.

seus principais resultados sejam contemplados por meio das publicações indexadas incluídas nesta revisão.

Conforme mencionado na Seção 3.5, outra limitação potencial decorre da escolha das bases consultadas. Embora tenham sido priorizadas quatro repositórios com ampla cobertura, optou-se por não incluir bases como *Web of Science* e Elsevier. Para mitigar o risco de omissão de estudos relevantes que pudessem estar indexados nessas e em outras fontes, foi realizada uma busca complementar no Google Acadêmico, o que amplia a cobertura, mas não elimina completamente a possibilidade de perdas.

Por fim, embora o processo de triagem e seleção tenha sido conduzido por três revisores, é possível que interpretações individuais tenham influenciado decisões pontuais. Para mitigar esse viés, divergências foram discutidas coletivamente até que se alcançasse consenso, assegurando maior consistência ao processo de seleção.

4 RESULTADOS E DISCUSSÕES

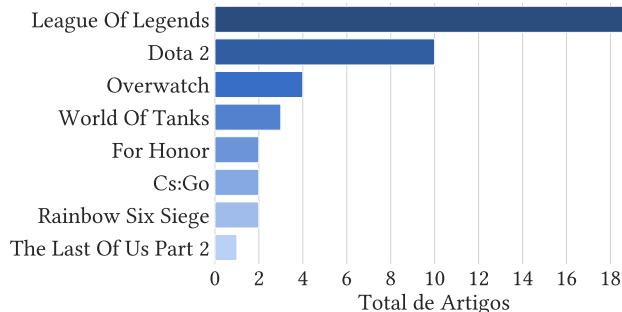
Esta seção apresenta os resultados obtidos após a aplicação das etapas de filtragem e triagem descritas na Seção 3. A listagem completa dos registros selecionados encontra-se na Tabela 2, contendo título, autores, ano de publicação e número de citações (coletado durante o período de realização da revisão de escopo).

Inicialmente, são exploradas características gerais dos estudos selecionados, incluindo a distribuição temporal das publicações, a recorrência de autores e outros aspectos relacionados aos metadados. Em seguida, são discutidas as três questões de pesquisa que guiam o trabalho, abordando os jogos e plataformas mais investigados, os métodos utilizados para detecção do discurso de ódio e as principais propostas de intervenção identificadas na literatura. Por fim, temos as lacunas e oportunidades de pesquisa identificadas.

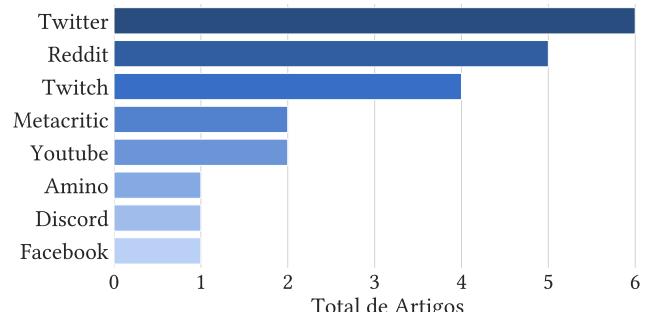
Além disso, para facilitar a síntese e organização dos achados desta revisão de escopo, ao efetuar essa análise, os estudos foram categorizados tematicamente em relação a cada uma das questões de pesquisa, conforme ilustrado na Figura 4, e explicado em suas seções correspondentes. Esta divisão hierárquica facilita a compreensão do panorama obtido e permite visualizar, de forma integrada, os principais ambientes analisados, as abordagens utilizadas para detecção de toxicidade e as propostas de intervenção discutidas na literatura, possibilitando identificar padrões, sobreposições e lacunas nas investigações.

Tabela 2: Sumário dos estudos selecionados após o processo de triagem.

Ref.	Titulo	Autor(es)	Ano	Citações
[73]	Deception in Video Games: Examining Varieties of Griefing	Rubim e Camm	2013	52
[9]	STFU NOOB! Predicting Crowdsourced Decisions on Toxic Behavior in Online Games	Blackburn e Kwak	2014	242
[4]	Automatic Analysis and Identification of Verbal Aggression and Abusive Behaviors For Online Social Games	Balci e Salah	2014	94
[47]	Linguistic Analysis of Toxic Behavior in an Online Video Game	Kwak e Blackburn	2015	143
[48]	Exploring Cyberbullying and Other Toxic Behavior in Team Competition Online Games	Kwak <i>et al.</i>	2015	425
[52]	Toxicity Detection in Multiplayer Online Games	Martens <i>et al.</i>	2015	154
[7]	Jogo Sujo: Violência Verbal e Liberdade de Expressão nos Games	Bernardes <i>et al.</i>	2016	1
[37]	Why do Teammates Hate Me? Cross-cultural Tensions and Social Dynamics in Online Games	Jiang e Yarosh	2016	4
[31]	Calm down buddy! It's just a game: Behavioral Patterns Observed Among Teamwork MMO Participants in WARGAMING's World of Tanks	Esmaeili e Woods	2016	18
[58]	Studying Toxic Behavior Influence and Player Chat in an Online Video Game	Neto <i>et al.</i>	2017	105
[17]	Hate is not Binary: Studying Abusive Behavior of #GamerGate on Twitter	Chatzakou <i>et al.</i>	2017	109
[25]	Relating Conversational Topics and Toxic Behavior Effects in a MOBA Game	Neto e Becker	2018	66
[78]	"A pure meritocracy blind to identity": Exploring the Online Responses to All-Female Teams in Reddit	Siuitila e Havaste	2018	47
[54]	Online Abuse and Age in Dota 2	Mattinen e Macey	2018	37
[26]	Front Line: Machismo nos Campos de Justiça do League of Legends	Souza e Rost	2019	9
[76]	Exploring the Relationship Between Game Content and Culture-based Toxicity	Sengün <i>et al.</i>	2019	37
[21]	Cyberbullying Detection for Online Games Chat Logs Using Deep Learning	Cornel <i>et al.</i>	2019	20
[20]	For Whom the Gamer Trolls: A Study of Trolling Interactions in the Online Gaming Context	Cook <i>et al.</i>	2020	65
[77]	Viral Vitriol: Predictors and Contagion of Online Toxicity in World of Tanks	Shen <i>et al.</i>	2020	123
[28]	Assessing Toxic Behaviour in Dead by Daylight Perceptions and Factors of Toxicity According to the Game's Official Subreddit Contributors	Deslauriers <i>et al.</i>	2020	31
[55]	Toxic Teammates or Obscene Opponents? Influences of Cooperation and Competition on Hostility Between Teammates and Opponents in an Online Game	McLean <i>et al.</i>	2020	28
[46]	Toxic Behaviors in Team-Based Competitive Gaming: The Case of League of Legends	Kou	2020	175
[35]	I'm just trolling: The Role of Normative Beliefs in Aggressive Behaviour in Online Gaming	Bruce e Neill	2020	190
[40]	Towards a Unified Theory of Toxic Behavior in Video Games	Kordyaka <i>et al.</i>	2020	168
[83]	See No Evil, Hear No Evil, Speak No Evil: How Collegiate Players Define, Experience and Cope with Toxicity	Türkay <i>et al.</i>	2020	143
[50]	"Who are the Cyka Blyat?" How Empathy, Impulsivity, and Motivations to Play Predict Aggressive Behaviors in Multiplayer Online Games	Lemerrier-Dugarin <i>et al.</i>	2021	42
[42]	Curing Toxicity – A Multi-method Approach	Kordyaka e Kruse	2021	0
[36]	The Mind Games Have Already Started: An In-Depth Examination of Trash Talking in Counter-Strike: Global Offensive Esports Using Practice Theory	Irwin <i>et al.</i>	2021	16
[41]	Exploring the Dark Side of Multiplayer Online Games: The Relationship Between Contact Experiences and Sexism	Kordyaka <i>et al.</i>	2022	4
[70]	Feeling Good and In Control: In-Game Tools to Support Targets of Toxicity	Reid <i>et al.</i>	2022	40
[69]	'Bad Vibrations': Sensing Toxicity From In-Game Audio Features	Reid <i>et al.</i>	2022	38
[56]	A Campanha de Ódio contra The Last of Us Part II	Melo e Pimentel	2022	3
[45]	Exploring the Relationship Between Offline Cultural Environments and Toxic Behavior Tendencies in Multiplayer Online Games	Kordyaka <i>et al.</i>	2023	34
[44]	The Saliency of Dispositions: Personality Traits, Anger, and Aggression as Antecedents of Toxicity in Multiplayer Online Battle Arena Games	Kordyaka <i>et al.</i>	2023	0
[84]	Unveiling Identity Biases in Toxicity Detection: A Game-Focused Dataset and Reactivity Analysis Approach	Dorpe <i>et al.</i>	2023	4
[57]	Toxic Play: Examining the Issue of Hate Within Gaming	Munn	2023	18
[43]	The Cycle of Toxicity: Exploring Relationships between Personality and Player Roles in Toxic Behavior in Multiplayer Online Battle Arena Games	Kordyaka <i>et al.</i>	2023	31
[68]	Suspecting Sarcasm: How League of Legends Players Dismiss Positive Communication in Toxic Environments	Poeller <i>et al.</i>	2023	22
[30]	Toxicity in Twitch Live Stream Chats: Towards Understanding the Impact of Gender, Size of Community and Game Genre	Dreier e Pirker	2023	3
[51]	"After All, They Don't Know Me" Exploring the Psychological Mechanisms of Toxic Behavior in Online Games	Liu e Agur	2023	45
[1]	Towards Detecting Contextual Real-Time Toxicity for In-Game Chat	Yang <i>et al.</i>	2023	7
[75]	"Creepy Towards My Avatar Body, Creepy Towards My Body": How Women Experience and Manage Harassment Risks in Social Virtual Reality	Schulenberg <i>et al.</i>	2023	51
[82]	"Pro Gamers" & Cyberbullying: Workplace Bullying & Sexual Harassment in Professional Video Gaming	Trudgett-Klose e McLinton	2024	14
[72]	Quando as Avaliações Viram Bombas: Explorando a Dinâmica do Review Bombing nos Jogos no Metacritic	Ribeiro <i>et al.</i>	2024	0
[2]	Game On, Hate Off: A Study of Toxicity in Online Multiplayer Environments	Yang <i>et al.</i>	2024	5
[39]	Challenges in Moderating Disruptive Player Behavior in Online Competitive Action Games	Kocielnik <i>et al.</i>	2024	7
[49]	Traumatizing or Just Annoying? Unveiling the Spectrum of Gamer Toxicity in the StarCraft II Community	Laato <i>et al.</i>	2024	10
[86]	How To Tame a Toxic Player? A Systematic Literature Review on Intervention Systems for Toxic Behaviors in Online Video Games	Wijkstra <i>et al.</i>	2024	14
[38]	Toxicity or Prosociality?: Civic Value and Gaming Citizenship in Competitive Video Game Communities	Kim e Ortiz	2024	2



(a) Análise dos jogos mais abordados



(b) Análise das plataformas mais abordados

Figura 3: Análise dos ambientes explorados sobre toxicidade e ódio em jogos.

4.1 Características gerais

No conjunto de estudos analisados (Tabela 2), destacam-se autores com presença recorrente e impacto relevante na literatura. Bas-tian Kordyaka é o nome mais frequente, com sete publicações selecionadas [40–45, 49], que somam 247 citações. Samuli Laato aparece em cinco trabalhos [41, 43–45, 49], com 79 citações, enquanto Regan L. Mandryk contribui com quatro artigos [68–70, 86], totalizando 114 citações. Jeremy Blackburn, também com quatro trabalhos selecionados [9, 17, 47, 48], se destaca pelo impacto expressivo: seus estudos acumulam 919 citações, o maior número observado, o que sugere elevado reconhecimento e contribuição para a área.

A frequência desses nomes, associada ao volume de citações, indica que os procedimentos de busca e seleção empregados foram eficazes em identificar produções científicas representativas do campo. Observa-se, ainda, a formação de redes de pesquisa consolidadas, exemplificada pela colaboração entre Laato e Kordyaka, que assinam conjuntamente todos os trabalhos em que Laato figura como autor [41, 43–45, 49].

Somado a isso, a Figura 2 apresenta a distribuição temporal dos estudos selecionados, revelando uma concentração significativa de publicações a partir de 2020. Esse dado pode refletir um crescimento no interesse da comunidade científica pelas manifestações de ódio em ambientes digitais e dentre eles, os jogos, possivelmente impulsionado por transformações sociais, culturais e tecnológicas ocorridas nos últimos anos [14].

Durante a pandemia de COVID-19, por exemplo, as restrições de mobilidade e isolamento social levaram muitas pessoas a recorrer a meios digitais tanto para manutenção de vínculos afetivos quanto para a busca de entretenimento, de forma que os jogos digitais passaram a desempenhar um papel ainda mais central no cotidiano, promovendo a interação social mediada por tecnologia e expandindo significativamente o volume e a diversidade dos usuários nesses espaços [5, 61]. Esse crescimento pode ter contribuído para o aumento da visibilidade e, possivelmente, da incidência de comportamentos tóxicos, à medida que novos públicos interagiam com culturas e dinâmicas já estabelecidas nas comunidades de jogadores. Dessa forma, a percepção desse fenômeno, tanto por jogadores quanto por pesquisadores, pode ter sido um dos fatores que motivaram o crescimento da produção acadêmica sobre o tema nos anos seguintes.

É relevante destacar, também, como os estudos revisados abordam a relação entre os conceitos de discurso de ódio e toxicidade.

A análise realizada indica que, na maioria dos trabalhos que definem esses termos, a toxicidade é utilizada como um conceito mais abrangente, englobando diferentes comportamentos nocivos nos ambientes de jogo – entre eles, o discurso de ódio. Essa compreensão mais ampla é observada em trabalhos como Blackburn e Kwak [2014], Kordyaka *et al.* [2020], Kou [2020], entre outros.

Diante desse cenário, torna-se essencial compreender de forma estruturada como o discurso de ódio em jogos digitais tem sido estudado. As próximas seções organizam os dados obtidos, respondendo às perguntas de pesquisa que guiam esta revisão.

4.2 QP1: Ambientes e contextos analisados

A literatura sobre discurso de ódio em jogos digitais pode ser organizada em dois grandes eixos de análise em relação ao ambiente onde ocorrem: (i) os próprios jogos, enquanto espaços de interação direta entre jogadores, e (ii) as plataformas digitais que orbitam esse universo, como fóruns, redes sociais e serviços de streaming. Ambos os eixos estão representados na seção de Ambientes da Figura 4.

O primeiro eixo, com 37 artigos, correspondente aos jogos como ilustrado na Figura 4, foca nas interações que ocorrem dentro dos próprios títulos. A Figura 3a ilustra os jogos mais analisados, que são, em sua maioria, online, multijogador e competitivos – sete dos oito títulos mais abordados compartilham esse perfil. Estudos como Blackburn e Kawk [2014] e Kou [2020] reforçam que a competição é uma das principais causas da alta toxicidade nesses jogos. Elas evidenciam que a natureza competitiva dos jogos online cria um ambiente propício para o surgimento e a disseminação de comportamentos tóxicos, frequentemente desencadeados por frustração, expectativas não atendidas e a pressão por resultados.

Entre esses jogos, destaca-se o *League of Legends* (LoL), que aparece como o título mais investigado (19 artigos), sendo tema de quase metade dos artigos revisados. Trata-se de um jogo do gênero MOBA, com uma base massiva de jogadores e amplamente reconhecido por sua comunidade marcada por comportamentos tóxicos. Além disso, outro fator que incentivou estudos sobre o jogo foi o Tribunal *League of Legends* [63], uma iniciativa oficial da desenvolvedora Riot Games, que disponibilizava publicamente casos de denúncias de conduta inadequada para julgamento pela própria comunidade. Essa ferramenta, ativa até 2014, oferecia acesso estruturado as evidências de toxicidade, como registros de partidas e

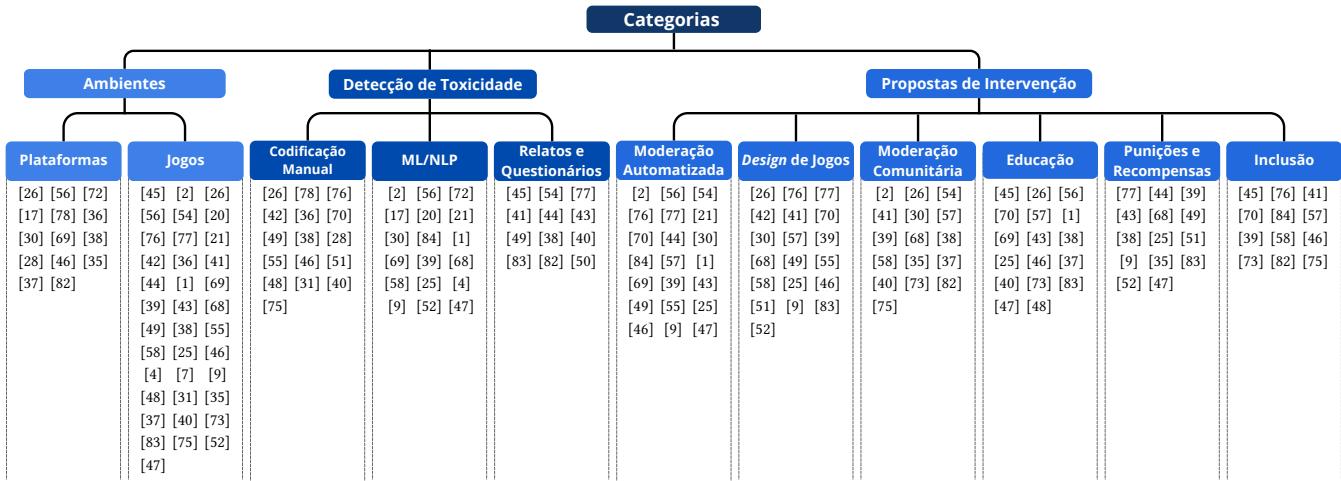


Figura 4: Categorização temática dos estudos selecionados nesta revisão de escopo segundo ambientes investigados, métodos de detecção de toxicidade e propostas de intervenção. Cada categoria principal é subdividida em subgrupos de acordo com características específicas dos estudos.

transcrições de chat, o que facilitou significativamente a coleta de dados por pesquisadores. Como resultado, o Tribunal foi amplamente utilizado como fonte de dados em diversos estudos [9, 25, 47, 48, 58].

Uma exceção a ser observada entre os jogos mais analisados é *The Last of Us Part II*, da desenvolvedora *Naughty Dog*, o único título da lista com jogabilidade para um jogador (*singleplayer*) e narrativa linear. Diferentemente dos demais, sua presença nos estudos selecionados não se deve a interações entre jogadores durante a partida, mas sim à reação odiosa massiva da parte da comunidade *gamer* em plataformas como Metacritic, Reddit¹¹ e YouTube¹². Esse jogo se tornou um caso emblemático de *review bombing*, sobretudo contra suas personagens femininas, como Abby e Ellie, frequentemente alvo de discursos misóginos e transfóbicos [56, 72], evidenciando como jogos altamente narrativos também podem catalisar discursos de ódio, especialmente quando desconstroem estereótipos tradicionais de gênero, desafiando expectativas de parte do público.

O segundo eixo, composto por 14 artigos e representado como plataformas na Figura 4, foca nas plataformas digitais que orbitam o ecossistema dos jogos digitais. Esses espaços, como redes sociais, fóruns e plataformas de *streaming*, que conforme mencionado anteriormente, podem funcionar como extensões do ambiente de jogo. A Figura 3b mostra que, embora esses ambientes sejam menos estudados que os próprios jogos, ainda representam um foco relevante de pesquisa [36, 82]. Entre as plataformas mais frequentemente analisadas nos estudos sobre discurso de ódio no contexto dos jogos digitais, destacam-se: Twitter¹³ (6 artigos), Reddit (5 artigos) e Twitch¹⁴ (4 artigos). Twitter e Reddit já possuem um histórico consolidado na literatura sobre discurso de ódio [16, 53, 65, 71], o que favorece sua adoção também em pesquisas voltadas ao universo dos jogos digitais. Outro fator que influencia nos estudos sobre essas redes são suas APIs – antigamente gratuitas e acessíveis, o que

facilitava a coleta de grandes volumes de dados textuais –, sendo utilizadas em artigos como Chatzakou *et al.* [2017] e Kou [2020].

Neste contexto, destaca-se, entre os episódios analisados, o *GamerGate*, frequentemente citado como marco na discussão sobre cultura *gamer* e discurso de ódio online [17]. Além desses casos amplos, estudos como Siutila e Havaste [2018], Deslauriers *et al.* [2020] e Kou [2020], também se voltam para nichos mais específicos, como os fóruns do Reddit, também conhecidos como subreddits, dedicados a jogos individuais, por exemplo, *r/leagueoflegends*¹⁵, *r/GlobalOffensive*¹⁶ e *r/deadbydaylight*¹⁷. Esses fóruns funcionam como micro comunidades, onde práticas tóxicas e discursos odiosos podem ser observados no dia-a-dia da interação entre jogadores.

Visando entender melhor quais tipos de dados são utilizados na análise do discurso de ódio, foi realizada uma categorização manual dos objetos de estudo e fonte de dados empregados pelos estudos revisados (Figura 5). Foram identificadas cinco categorias principais: (i) comentários, que incluem mensagens textuais extraídas de fóruns, chats e redes sociais; (ii) usuários, referentes a informações sobre os perfis ou comportamentos dos indivíduos criadores e alvos deste tipo de conteúdo; (iii) partidas, associado a dados relacionados ao contexto das sessões de jogo, como duração, resultados e eventos ocorridos; (iv) participantes, no caso de estudos experimentais ou observacionais com jogadores em ambientes controlados; (v) entrevistas, quando utilizadas para captar percepções, experiências e relatos subjetivos.

Analizando a Figura 5a, que apresenta os objetos de estudos mais utilizados pelos pesquisadores, é possível identificar duas vertentes distintas nas abordagens dos estudos. A primeira é mais impersonal e quantitativa. Pesquisas como as de Balci e Salah [2014], Kwak e Blackburn [2015] e Yang *et al.* [2024] trabalham com grandes volumes de dados objetivos, como comentários, partidas, usuários e posts. Esses trabalhos utilizam algoritmos de aprendizado de

¹¹Acessível em: <https://www.reddit.com/>

¹²Acessível em: <https://www.youtube.com/>

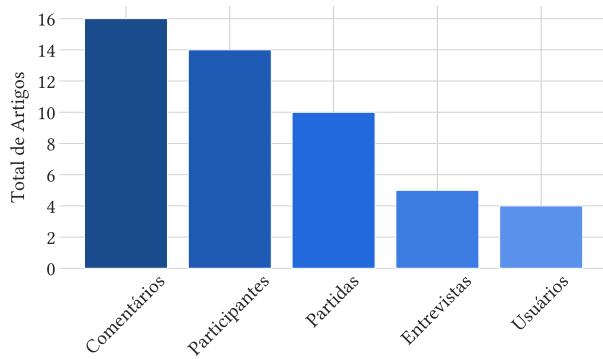
¹³Acessível em: <https://x.com/>

¹⁴Acessível em: <https://twitch.tv/>

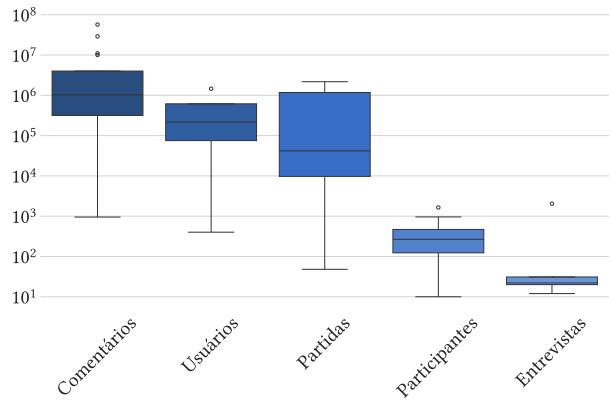
¹⁵Acessível em: <https://www.reddit.com/r/leagueoflegends/>

¹⁶Acessível em: <https://www.reddit.com/r/GlobalOffensive/>

¹⁷Acessível em: <https://www.reddit.com/r/deadbydaylight/>



(a) Objetos de estudo mais utilizados.



(b) Distribuição da quantidade de dados por abordagem.

Figura 5: Análise dos objetos de estudos e fontes de dados utilizados pelos estudos.

máquina e técnicas de processamento de linguagem natural (NLP), o que permite análises em larga escala, às vezes com menor sensibilidade às nuances contextuais. Já a segunda vertente é mais qualitativa e subjetiva. Estudos como os de Kim e Ortiz [2024], Kordyaka *et al.* [2022], Laato *et al.* [2024] e Liu e Agur [2023] focam nas experiências humanas, utilizando dados coletados por meio de entrevistas e interações com participantes. Esse tipo de abordagem permite compreender com mais profundidade os impactos sociais e emocionais do discurso de ódio, aspectos que dificilmente seriam captados apenas por métricas textuais ou números.

Ao analisar a quantidade de dados utilizados em cada categoria, conforme representado na Figura 5b, observa-se que os estudos nos quais são utilizados dados quantitativos tendem a operar com volumes significativamente maiores. A mediana para o número de comentários, por exemplo, é de aproximadamente 1 milhão, com alguns estudos como Kwak e Blackburn [2015] com mais de 50 milhões de comentários. Por outro lado, estudos baseados em entrevistas e dados fornecidos diretamente por participantes operam com volumes bem menores, geralmente entre dezenas e centenas de instâncias [38, 40, 54, 75]. Essa diferença reflete, além de estratégias metodológicas distintas, desigualdades nas formas de coleta e disponibilidade dos dados. Estudos quantitativos frequentemente se beneficiam de coletas massivas por meio de APIs e técnicas de *Web scraping*. Já os qualitativos, por dependerem de interações humanas, enfrentam maiores limitações em termos de escala, mas costumam ganhar em relação à qualidade dos relatos e especificidade dos dados levantados. Um exemplo atípico é o estudo de Poeller *et al.* [2023], que apresenta o maior número de participantes entre os qualitativos, com 959 respostas.

4.3 QP2: Técnicas de detecção utilizadas

Conforme representado na parte de detecção de toxicidade na categorização hierárquica criada, as técnicas e abordagens que têm sido utilizadas para a detecção do discurso de ódio dentro do ambientes de jogos nos estudos analisados se distribuem principalmente entre três categorias centrais: métodos computacionais, técnicas qualitativas interpretativas e instrumentos baseados em autorrelato.

A primeira e mais frequente forma de detecção de toxicidade envolve modelos de *Machine Learning* (ML) e NLP, utilizados em 18 estudos. Esses trabalhos exploram uma variedade de métodos supervisionados e não supervisionados, incluindo redes neurais profundas, como Redes Neurais Convolucionais (*Convolutional Neural Networks*, CNN) [21], *Bidirectional Encoder Representation Transformer* (BERT) [29] e *Naïve Bayes Point Machine* [21]; modelos clássicos, como Máquinas de Vetores de Suporte (*Support Vector Machines*, SVM) [8] e *Random Forests* [12]; além de ferramentas específicas, como Perspective API¹⁸, ToxBuster¹⁹, SentiStrength²⁰ e LIWC²¹ (*Linguistic Inquiry and Word Count*) [2, 17, 21]. Algumas abordagens incorporam ainda a análise de N-gramas, modelagem de tópicos com LDA – do inglês, *Latent Dirichlet Allocation* –, características semânticas e cronêmicas, além de dados contextuais, como metadados dos jogadores e históricos de chat [1, 58, 72]. Também são utilizadas técnicas multimodais e análises de reatividade, como no estudo que avalia o viés de identidade na detecção de toxicidade, por meio da construção de conjuntos de dados sintéticos e da medição da sensibilidade dos modelos a termos ambíguos [84]. Esses métodos demonstram alto potencial de escalabilidade e objetividade, embora enfrentem desafios como o viés algorítmico e a dificuldade em capturar ironias, sarcasmos e contextos sutis.

A segunda técnica de detecção destacada na Figura 4 refere-se à técnicas qualitativas de detecção de toxicidade, presente em 16 artigos. Essa categoria abrange métodos mais interpretativos, como codificação manual, a análise de conteúdo, o *close reading*²², a análise temática reflexiva e indutiva, além de observações etnográficas [38, 49, 78]. Em muitos casos, pesquisadores ou avaliadores treinados analisam entrevistas, interações em *chats*, vídeos de sessões de jogo ou dados observacionais, buscando interpretar o contexto e os significados subjetivos expressos nas comunicações [31, 38, 70]. Diversos estudos também combinam essas

¹⁸Disponível em: <https://www.perspectiveapi.com/>¹⁹Disponível em: <https://github.com/ubisoft/ubisoft-laforge-toxbuster>²⁰Disponível em: <http://sentistrength.wlv.ac.uk/>²¹Disponível em: <https://www.liwc.app/>²²O termo *close reading* se refere à leitura minuciosa e aprofundada dos textos, voltada para a identificação de nuances linguísticas, estruturas discursivas e significados implícitos [11].

análises com escalas psicométricas ou classificações manuais realizadas por jogadores, ou moderadores, como no sistema Tribunal, mencionado anteriormente, que media julgamentos coletivos com base em evidências comportamentais [42, 55]. Embora esse tipo de abordagem seja eficaz na identificação de nuances e contextos complexos, ela apresenta limitações em termos de escalabilidade.

Por fim, a terceira abordagem de relatos e questionários representada na Figura 4 envolve métodos baseados em questionários e autorrelato de forma mais indireta com a toxicidade, utilizados em 12 estudos. Essa abordagem é frequentemente combinada com entrevistas qualitativas e análise temática, com o objetivo de captar as percepções subjetivas dos próprios jogadores sobre suas experiências com discurso de ódio. Os instrumentos utilizados incluem questionários online, escalas validadas, como a UTTB – do inglês, *Unified Theory of Toxic Behavior* –, que avalia o nível de compreensão e aceitação de comportamentos tóxicos [45]. Além disso, são empregadas outras escalas direcionadas à avaliação de atitudes relacionadas ao sexismo, empatia e normas sociais em ambientes digitais [41, 50, 83]. Esses métodos permitem um entendimento aprofundado das vivências individuais e dos impactos emocionais do discurso de ódio, sendo particularmente úteis para identificar formas mais veladas e subjetivas de agressão.

4.4 QP3: Propostas de mitigação discutidas

Além de estudar o fenômeno, entender as motivações e canais por onde a toxicidade e discurso de ódio acontece nos jogos, vários estudos também propõem meios para combate desta prática. Com base na secção de propostas de intervenção da Figura 4, os artigos analisados abordam diversas estratégias de mitigação do discurso de ódio em jogos digitais. As propostas foram agrupadas em seis grandes categorias temáticas, conforme descrito a seguir.

A primeira e mais recorrente categoria da Figura 4 é a de moderação automatizada e ferramentas técnicas, contemplando 21 artigos. Essas propostas sugerem a utilização de inteligência artificial, aprendizado de máquina e sistemas híbridos para detectar e intervir sobre comportamentos tóxicos em tempo real [1, 21, 49]. São apresentadas soluções como *benchmarks* de viés voltados para jogos, que identificam e mitigam falsos positivos em termos relacionados à identidade – como raça, gênero e orientação sexual –, modelos preditivos, ferramentas de moderação de chat, *shadow bans* – em sua tradução livre, bloqueios silenciosos –, extensões de navegador e penalidades automatizadas [9, 43, 69, 84]. Diversos trabalhos propõem abordagens proativas, como alertas visuais a jogadores potencialmente tóxicos e bloqueios temporários de comunicação em momentos de conflito, visando reduzir a carga da moderação humana e atuar preventivamente sobre surtos de toxicidade [1, 77].

A segunda categoria refere-se ao *design* de jogos e mecânicas pró-sociais, abordada em 19 artigos. Ela destaca o papel do próprio ambiente e estrutura dos jogos na prevenção de comportamentos tóxicos. As estratégias incluem ajustes no sistema de pareamento de partidas, incentivos à formação de comunidades colaborativas, *design* de funcionalidades voltadas ao reforço de empatia, mecanismos de autorregulação e ferramentas emocionais [38, 41, 52, 76]. Também foram identificadas propostas centradas em ajustes de *design*, como botões de reconhecimento – “gostei” ou “elogiar” –

e redução do anonimato, promovendo a humanização entre os jogadores [42, 51].

A terceira categoria da Figura 4 é a moderação comunitária, que compreende 16 artigos e reúne iniciativas que ampliam a participação dos próprios jogadores na identificação e resposta ao discurso de ódio. Os estudos sugerem mecanismos de denúncia mais eficientes, como os do tipo acumulativo, e pré-marcação de conteúdos tóxicos por humanos [38, 69]. Também propõem o fortalecimento da moderação comunitária voluntária com revisões coletivas – como no sistema Tribunal –, além da adaptação de sistemas de denúncia a diferentes culturas e normas locais [37, 48].

A quarta categoria apresentada na Figura 4 trata de educação e conscientização, sendo discutida em 17 artigos. Nela, as propostas vão desde campanhas educativas e programas formais, como o *Game Over Hate* – uma iniciativa que visa combater o assédio em jogos por meio de *workshops*, sensibilização dos jogadores e promoção de ambientes mais inclusivos – até a integração de normas de conduta e comunicação saudável no ecossistema dos jogos [57, 83]. Há também foco no discurso de ódio como forma de *cyberbullying* e na promoção de ambientes de jogo mais seguros e inclusivos [82]. Estudos como Hilvert-Bruce e Neill [2020] sugerem a atuação conjunta de empresas, influenciadores e comunidade para fomentar normas coletivas e práticas empáticas, bem como a educação sobre temas como sexismo, responsabilidade individual e impacto das palavras nos ambientes virtuais. A quinta categoria evidenciada na Figura 4 aborda o uso de recompensas e punições, com 14 artigos. As propostas associam comportamentos positivos a recompensas *in-game* – como progressão, bônus ou melhorias no *matchmaking*²³ –, ao mesmo tempo em que preveem punições graduais e pedagógicas, como silenciamentos, suspensões e bloqueios automáticos [43, 52, 70]. Destacam-se ainda iniciativas que personalizam as respostas com base na personalidade ou histórico do jogador, além do estímulo a intervenções positivas de observadores e medidas para reabilitação de usuários reincidentes [38, 44].

Por fim, a sexta categoria da Figura 4 trata de adaptações culturais e inclusão, presente em 12 artigos. Ela aborda intervenções sensíveis às características socioculturais dos jogadores. Isso inclui o combate a estereótipos no conteúdo dos jogos, a inclusão de comunidades diversas no processo de *design*, e a adaptação das ferramentas de moderação à realidade geográfica e cultural dos servidores [45, 76]. Também são discutidas estratégias voltadas à equidade, como o enfrentamento de estruturas discriminatórias, a educação sobre preconceitos de gênero e o desenvolvimento de ferramentas contextualmente sensíveis [45, 75]. As propostas buscam promover uma representação significativa nos espaços de jogo e tornar as ações de moderação mais justas e inclusivas.

4.5 Oportunidades e Lacunas de Pesquisa

Os resultados sintetizam algumas lacunas recorrentes na literatura, sendo a primeira delas a escassez de conjunto de dados abertos e acessíveis. Dos 49 artigos analisados, apenas quatro disponibilizam seus conjuntos de dados, e somente o estudo de Poeller *et al.* [2023] oferece acesso totalmente público. Os demais conjuntos [25, 39, 58] exigem solicitações diretas aos autores. Essa limitação compromete

²³O termo *matchmaking* se refere ao processo de pareamento de jogadores, conectando dois ou mais participantes para compor uma partida em um jogo [15].

a replicabilidade dos estudos e dificulta a criação de *benchmarks* consistentes para o discurso de ódio, restringindo o progresso colaborativo da área, assim como a reproduzibilidade dos resultados obtidos.

Além disso, há uma carência na validação prática das propostas de intervenção: embora 38 dos 49 artigos (77.5%) apresentem sugestões concretas, nenhum deles implementa ou testa essas propostas em ambientes reais. Isso revela um descompasso entre a formulação teórica das medidas e sua verificação empírica. Trabalhos como o de Van Dorpe *et al.* [2023], que propõem metodologias experimentais para avaliar vieses em classificadores, podem servir de modelo para desenvolver protocolos de validação em contextos controlados ou aplicados. Da mesma forma, o estudo de Wijkstra [2024] evidencia que a maioria das propostas analisadas são de sistemas de moderação ainda não testados, enquanto apenas uma minoria corresponde à avaliação de sistemas já existentes, reforçando as observações aqui apresentadas.

Além disso, há plataformas centrais para a experiência *gamer* que permanecem pouco exploradas na literatura sobre discurso de ódio. Um exemplo notável é a Steam²⁴, atualmente a maior plataforma de distribuição digital de jogos do mundo, com ampla base de usuários e funcionalidades sociais como fóruns, avaliações, grupos e perfis públicos. Apesar de sua centralidade no ecossistema *gamer*, não foi identificada em nenhum dos estudos revisados. O mesmo vale para outras plataformas sub-exploradas como Facebook²⁵, Discord²⁶, Amino²⁷ e Metacritic⁵, que embora pontualmente mencionadas, ainda carecem de estudos sistemáticos.

Por fim, observa-se uma baixa representatividade da produção científica nacional na área. Dos 49 trabalhos completos analisados, apenas 1 (2%) analisou diretamente o contexto brasileiro [26]. Mesmo com a relevância do país nessa indústria, esse dado evidencia a sub-representação dos jogadores brasileiros nos debates internacionais sobre discurso de ódio em jogos digitais. Essa lacuna pode limitar a compreensão de dinâmicas específicas do contexto sociocultural brasileiro, bem como a formulação de estratégias localizadas de intervenção. Dada a relevância crescente do tema e a expressiva base de jogadores no país, há uma oportunidade importante para ampliar a participação da comunidade acadêmica nacional nesse campo de pesquisa.

5 CONSIDERAÇÕES FINAIS

Esta revisão de escopo mapeou sistematicamente a produção científica sobre toxicidade e discurso de ódio em jogos digitais, analisando 49 estudos publicados entre 2010 e 2025. O levantamento evidencia que a área tem se expandido de forma significativa, especialmente após 2020, acompanhando o crescimento da relevância social e cultural dos jogos como espaços de interação. Ainda assim, a pesquisa sobre o tema permanece marcada por fragmentações e lacunas que limitam sua consolidação.

Entre as contribuições deste trabalho está a sistematização de aspectos-chave da literatura, incluindo os ambientes mais investigados – com destaque para jogos multijogador e competitivos, como *League of Legends* e *DOTA 2* – e a análise de plataformas complementares, como fóruns, redes sociais e serviços de *streaming*, onde

²⁴Acessível em: <https://store.steampowered.com/>

²⁵Acessível em: <https://www.facebook.com/>

²⁶Acessível em: <https://www.discord.gg/>

²⁷Acessível em: <https://aminoapps.com/>

Twitter e Reddit se destacam. A revisão identifica três frentes para a detecção de discurso de ódio – modelos computacionais, análises qualitativas e instrumentos de autorrelato. Essa diversidade revela avanços importantes, além de limitações relacionadas à escalabilidade e à subjetividade. Por fim, a revisão identifica seis categorias predominantes de estratégias de mitigação, que vão de soluções técnicas e automatizadas a ações de conscientização e intervenções no design dos jogos.

Ao mesmo tempo, o estudo destaca desafios centrais que a área ainda enfrenta. Poucos trabalhos tornam seus conjuntos de dados acessíveis, dificultando a replicabilidade e a criação de *benchmarks* comparativos. A representatividade brasileira é limitada – apenas um estudo analisava o contexto brasileiro, apesar do país figurar entre os dez maiores mercados *gamer* do mundo. Além disso, nota-se um foco concentrado em certos ambientes (como MOBAs e redes sociais específicas), enquanto plataformas amplamente utilizadas, como a *Steam*, permanecem pouco exploradas.

O discurso de ódio em jogos digitais se configura como um fenômeno complexo, que transcende questões técnicas e envolve dimensões sociais, culturais e psicológicas. Ao reunir e organizar a literatura existente, esta revisão contribui para um entendimento mais integrado do campo e aponta caminhos para futuras investigações. Espera-se que este mapeamento sirva como referência para pesquisadores e como ponto de partida para iniciativas que promovam ambientes de jogo mais inclusivos, diversos e acolhedores para toda a comunidade de jogadores. Como continuidade deste estudo, propõe-se investigar contextos ainda pouco explorados, podendo citar o cenário nacional e outras plataformas relevantes, avaliar a efetividade de diferentes estratégias de mitigação e analisar como fatores culturais e de *design* de jogo influenciam a ocorrência e a percepção do discurso de ódio.

AGRADECIMENTOS

Este trabalho foi realizado com apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e do Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional, Tratamento e Disseminação de Informação (INCT-TILD-IAR) – Grant #408490/2024-1.

REFERÊNCIAS

- [1] Zachary Yang, Nicolas Grenon-Godbout, and Reihaneh Rabbany. 2023. Towards Detecting Contextual Real-Time Toxicity for In-Game Chat. In *Find. Assoc. Comput. Linguist.: EMNLP*. Association for Computational Linguistics (ACL), Singapore, 9894–9906. <https://doi.org/10.18653/v1/2023.findings-emnlp.663>
- [2] Zachary Yang, Nicolas Grenon-Godbout, and Reihaneh Rabbany. 2024. Game on, Hate off: A Study of Toxicity in Online Multiplayer Environments. *ACM Games* 2, 2 (2024), 1–13. <https://doi.org/10.1145/3675805>
- [3] Hilary Arksey and Lisa O’Malley. 2005. Scoping Studies: Towards a Methodological Framework. *International Journal of Social Research Methodology* 8, 1 (2005), 19–32. <https://doi.org/10.1080/1364557032000119616>
- [4] Koray Balci and Albert Ali Salah. 2014. Automatic Analysis and Identification of Verbal Aggression and Abusive Behaviors for Online Social Games. *Computers in Human Behavior* 53 (2014), 517–526. <https://doi.org/10.1016/j.chb.2014.10.025>
- [5] Matthew Barr and Alicia Copeland-Stewart. 2022. Playing Video Games During the COVID-19 Pandemic and Effects on Players’ Well-Being. *Games and Culture* 17, 1 (2022), 122–139. <https://doi.org/10.1177/15554120211017036>
- [6] Nicole A Beres, Julian Frommel, Elizabeth Reid, Regan L Mandryk, and Madison Klarkowski. 2021. Don’t You Know That You’re Toxic: Normalization of Toxicity

- in Online Gaming. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 438, 15 pages. <https://doi.org/10.1145/3411764.3445157>
- [7] Danilo Coronado Bernardes, Henrique Fernandes, and Victor José Henrique Bonaldi. 2016. Jogo Sujo: Violência Verbal e Liberdade de Expressão Nos Games. In *Proceedings of SBGames 2016*. Sociedade Brasileira de Computação (SBC), Porto Alegre, RS, Brasil, 1119–1122.
- [8] Christopher M. Bishop and Hugh Bishop. 2024. *Deep Learning: Foundations and Concepts*. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-031-45468-4>
- [9] Jeremy Blackburn and Haewoon Kwak. 2014. STFU NOOB! predicting crowd-sourced 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. <https://doi.org/10.1145/2566486.2567987>
- [10] CNN Brasil. 2023. E-sports: O que são e por que estão cada vez mais populares? <https://www.cnnbrasil.com.br/esportes/outros-esportes/e-sports/>
- [11] Clare Braun. 2022. "What Is Close Reading?" || Definition and Strategies | College of Liberal Arts | Oregon State University. <https://liberalarts.oregonstate.edu/wlf/what-close-reading-definition-and-strategies>
- [12] Leo Breiman. 2001. Random Forests. *Machine Learning* 45, 1 (2001), 5–32. <https://doi.org/10.1023/A:1010933404324>
- [13] Equipe Red Bull. 2018. Conheça as Iniciativas Women in Gaming e #MyGameMyName. <https://www.redbull.com/br-pt/iniciativas-mygamemyname-womeningaming>.
- [14] Stephanie Bührer, Kevin Koban, and Jörg Matthes. 2024. The WWW of digital hate perpetration: What, who, and why? A scoping review. *Computers in Human Behavior* 159 (2024), 108321. <https://doi.org/10.1016/j.chb.2024.108321>
- [15] Canaltech. 2022. O que é matchmaking? <https://canaltech.com.br/games/o-que-e-matchmaking-219302/>. Seção: Games. Acesso em 10 ago. 2025.
- [16] Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech. *Proceedings of the ACM on human-computer interaction* 1, CSCW (2017), 1–22.
- [17] Despoina Chatzakou, Nicolas Kourtellis, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Athena Vakali. 2017. Hate Is Not Binary: Studying Abusive Behavior of #GamerGate on Twitter. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media* (Prague, Czech Republic) (HT '17). Association for Computing Machinery, New York, NY, USA, 65–74. <https://doi.org/10.1145/3078714.3078721>
- [18] Redação Maria Claire. 2022. Entenda o gamergate: escândalo começou com divulgação de dados de desenvolvedora por ex-namorado. <https://revistamarieclaire.globo.com/Mulheres-do-Mundo/noticia/2015/05/entenda-o-gamergate-escandalo-comecou-com-divulgacao-de-dados-de-desenvolvedora-por-ex-namorado.html>
- [19] Mia Consalvo. 2012. Confronting Toxic Gamer Culture: A Challenge for Feminist Game Studies Scholars. *Ada: A Journal of Gender, New Media, and Technology* 1 (2012), 1–7. <https://doi.org/10.7264/N33X84KH> Acessado em 17 jul. 2025.
- [20] Christine Cook, Rianne Conijn, Juliette Schaafsma, and Marjolijn Antheunis. 2020. For Whom the Gamer Trolls: A Study of Trolling Interactions in the Online Gaming Context. *Journal of Computer-Mediated Communication* 24, 6 (2020), 293–318. <https://doi.org/10.1093/jcmc/zmz014>
- [21] James Albert Cornel, Carl Christian Pablo, Jan Arnold Marzan, Vince Julius Mercado, Bernie Fabito, Ramon Rodriguez, Manolito Jr. Octaviano, Nathaniel Oco, and Angelica De La Cruz. 2019. Cyberbullying Detection for Online Games Chat Logs Using Deep Learning. In *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*. IEEE, Laoag, Philippines, 1–5. <https://doi.org/10.1109/HNICEM48295.2019.9072811>
- [22] Mariana Coutinho. 2013. O que são 'trolls' e o que é 'trollagem'? <https://www.techtudo.com.br/noticias/2013/06/o-que-sao-trolls-e-o-que-e-trollagem.shtml>
- [23] Carolyn Cunningham. 2018. *Games Girls Play: Contexts of Girls and Video Games*. Lexington Books, United States of America.
- [24] Victor de Abreu. 2020. O que é griefing? Entenda prática proibida no LoL, CS:GO e mais jogos. <https://www.techtudo.com.br/noticias/2020/04/o-que-e-griefing-entenda-pratica-proibida-no-lol-csgo-e-mais-jogos-esports.shtml>
- [25] Joaquim Alvino da Mesquita Neto and Karin Becker. 2018. Relating Conversational Topics and Toxic Behavior Effects in a MOBA Game. *Entertainment Computing* 26 (2018), 10–29. <https://doi.org/10.1016/j.entcom.2017.12.004>
- [26] Bruna da Veiga de Souza and Laura Roncaglio Rost. 2019. Front Line: Machismo nos campos de justiça do League of Legends. In *Anais do XVIII SBGames – Simpósio Brasileiro de Jogos e Entretenimento Digital*. Sociedade Brasileira de Computação (SBC), Rio de Janeiro, RJ, Brasil, 729–736. <https://www.sbgames.org/sbgames2019/files/papers/CulturaFull/198156.pdf> Trilha Cultura – Full Papers.
- [27] Jan de Wit, Alicia van der Kraan, and Joep Theeuwes. 2020. Live Streams on Twitch Help Viewers Cope With Difficult Periods in Life. *Frontiers in Psychology* 11 (2020), 586975. <https://doi.org/10.3389/fpsyg.2020.586975>
- [28] Patrick Deslauriers, Laura Iseut Lafrance St-Martin, and Maude Bonenfant. 2020. Assessing Toxic Behaviour in Dead by Daylight: Perceptions and Factors of Toxicity According to the Game's Official Subreddit Contributors. *Game Studies* 20, 4 (2020). https://gamestudies.org/2004/articles/deslauriers_iseutlafrancemartain_bonenfant
- [29] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- [30] Lukas Dreier and Johanna Pirker. 2023. Toxicity in Twitch Live Stream Chats: Towards Understanding the Impact of Gender, Size of Community and Game Genre. In *2023 IEEE Conference on Games (CoG)*. IEEE, Boston, MA, USA, 1–4. <https://doi.org/10.1109/CoG57401.2023.10333159>
- [31] Human Esmaili and Peter Charles Woods. 2016. Calm Down Buddy! It's Just a Game: Behavioral Patterns Observed among Teamwork MMO Participants in WARGAMING's World of Tanks. In *2016 22nd International Conference on Virtual System & Multimedia (VSMM)*, Harold Thwaites, Sian Lun Lau, and Alonzo C. Addison (Eds.). IEEE, Kuala Lumpur, Malaysia, 1–11. <https://doi.org/10.1109/VSMM.2016.7863147>
- [32] Paula Fortuna and Sérgio Nunes. 2018. A survey on automatic detection of hate speech in text. *Comput. Surveys* 51, 4 (2018), 1–30. <https://doi.org/10.1145/3232676>
- [33] Ignacio Gagliardone, Danit Gal, Thiago Alves, and Martinez Sainz Martinez. 2015. *Countering online hate speech*. Unesco Publishing, Online.
- [34] Enrico Gandolfi, Richard E Ferdig, and Ilker Soyturk. 2023. Exploring the Learning Potential of Online Gaming Communities: An Application of the Game Communities of Inquiry Scale. *New Media & Society* 25, 6 (2023), 1374–1393. <https://doi.org/10.1177/14614448211027171>
- [35] Zorah Hilvert-Bruce and James T. Neill. 2020. I'm Just Trolling: The Role of Normative Beliefs in Aggressive Behaviour in Online Gaming. *Computers in Human Behavior* 102 (2020), 303–311. <https://doi.org/10.1016/j.chb.2019.09.003>
- [36] Sidney Irwin, Anjum Naweed, and Michele Lastella. 2021. The Mind Games Have Already Started: An in-Depth Examination of Trash Talking in Counter-Strike: Global Offensive Esports Using Practice Theory. *Journal of Gaming and Virtual Worlds* 13, 2 (2021), 173–194. https://doi.org/10.1386/JGVW_00035_1
- [37] Jialun Jiang and Svetlana Yarosh. 2016. Why Do Teammates Hate Me? Cross-Cultural Tensions and Social Dynamics in Online Games. In *CSCW '16 Companion: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing*. Association for Computing Machinery, San Francisco, CA, USA, 301–304. <https://doi.org/10.1145/2818052.2869087>
- [38] Jin Kim and Naishly Ortiz. 2024. Toxicity or Prosociality?: Civic Value and Gaming Citizenship in Competitive Video Game Communities. *Simulation and Gaming* 55, 6 (2024), 1057–1077. <https://doi.org/10.1177/10468781241277899>
- [39] Rafal Kocielnik, Zhuofang Li, Claudia Kann, Deshawn Sambrano, Jacob Morrier, Mitchell Linegar, Carly Taylor, Min Kim, Nabila Naqvie, Feri Soltani, Arman Dehpanah, Grant Cahill, Animashree Anandkumar, and R. Michael Alvarez. 2024. Challenges in Moderating Disruptive Player Behavior in Online Competitive Action Games. *Frontiers in Computer Science* 6 (2024), 1283735. <https://doi.org/10.3389/fcomp.2024.1283735>
- [40] Bastian Kordyaka, Katharina Jahn, and Bjoern Niehaves. 2020. Towards a Unified Theory of Toxic Behavior in Video Games. *Internet Research* 30, 4 (2020), 1081–1102. <https://doi.org/10.1108/INTR-08-2019-0343>
- [41] Bastian Kordyaka, Jeanine Krath, Samuli Laato, Solip Park, Katharina Jahn, and Björn Niehaves. 2022. Exploring the Dark Side of Multiplayer Online Games: The Relationship between Contact Experiences and Sexism. In *Proceedings of the 28th Americas Conference on Information Systems (AMCIS 2022)*. Association for Information Systems, Minneapolis, MN, USA, 1–10. <https://aisel.aisnet.org/amcis2022/vcc/vcc/3/> Track: Virtual Communities and Collaboration; Paper No. 1440.
- [42] Bastian Kordyaka and Björn Kruse. 2021. Curing Toxicity – A Multi-method Approach. In *Innovation Through Information Systems: Volume II (Lecture Notes in Information Systems and Organisation, Vol. 47)*, Frederik Ahlemann, Reinhard Schütte, and Stefan Stieglitz (Eds.). Springer, Cham, Switzerland, 340–346. https://doi.org/10.1007/978-3-030-86797-3_23
- [43] Bastian Kordyaka, Samuli Laato, Katharina Jahn, Juho Hamari, and Björn Niehaves. 2023. The Cycle of Toxicity: Exploring Relationships between Personality and Player Roles in Toxic Behavior in Multiplayer Online Battle Arena Games. *Proceedings of the ACM on Human-Computer Interaction* 7, CHI PLAY (2023), 611–641. <https://doi.org/10.1145/3611043>
- [44] Bastian Kordyaka, Samuli Laato, Sebastian Weber, and Gerhard Klassen. 2023. The Saliency of Dispositions: Personality Traits, Anger, and Aggression as Antecedents of Toxicity in Multiplayer Online Battle Arena Games. In *Proceedings of the 34th Australasian Conference on Information Systems (ACIS 2023)*. Association for Information Systems, Wellington, New Zealand. <https://aisel.aisnet.org/acis2023/129>

- [45] Bastian Kordyaka, Solip Park, Jeanine Krath, and Samuli Laato. 2023. Exploring the Relationship between Offline Cultural Environments and Toxic Behavior Tendencies in Multiplayer Online Games. *ACM Transactions on Social Computing* 6, 1–2, Article 3 (2023), 20 pages. <https://doi.org/10.1145/3580346>
- [46] Yubo Kou. 2020. Toxic Behaviors in Team-Based Competitive Gaming: The Case of League of Legends. In *CHI PLAY '20: Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Virtual Event (originally Ottawa, Canada)). Association for Computing Machinery, New York, NY, USA, 81–92. <https://doi.org/10.1145/3410404.3414243>
- [47] Haewoon Kwak and Jeremy Blackburn. 2015. Linguistic Analysis of Toxic Behavior in an Online Video Game. In *Social Informatics – SocInfo 2014 International Workshops, Barcelona, Spain, November 11, 2014, Revised Selected Papers*, Luca Maria Aiello and Daniel A. McFarland (Eds.), Lecture Notes in Computer Science, Vol. 8852. Springer, Cham, 209–217. https://doi.org/10.1007/978-3-319-15168-7_26
- [48] Haewoon Kwak, Jeremy Blackburn, and Seungyeop Han. 2015. Exploring Cyberbullying and Other Toxic Behavior in Team Competition Online Games. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) (CHI '15). Association for Computing Machinery, New York, NY, USA, 3739–3748. <https://doi.org/10.1145/2702123.2702529>
- [49] Samuli Laato, Bastian Kordyaka, and Juhu Hamari. 2024. Traumatizing or Just Annoying? Unveiling the Spectrum of Gamer Toxicity in the StarCraft II Community. In *CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 758, 18 pages. <https://doi.org/10.1145/3613904.3642137>
- [50] Maud Lemercier-Dugarin, Lucia Romo, Charles Tijus, and Oulmann Zerhouni. 2021. "who Are the Cyka Blyat?" How Empathy, Impulsivity, and Motivations to Play Predict Aggressive Behaviors in Multiplayer Online Games. *Cyberpsychology, Behavior, and Social Networking* 24, 1 (2021), 63–69. <https://doi.org/10.1089/cyber.2020.0041>
- [51] Yansheng Liu and Colin Agur. 2023. "After All, They Don't Know Me" Exploring the Psychological Mechanisms of Toxic Behavior in Online Games. *Games and Culture* 18, 5 (2023), 598–621. <https://doi.org/10.1177/15554120221115397>
- [52] Marcus Märtens, Siqi Shen, Alexandru Iosup, and Fernando Kuipers. 2015. Toxicity Detection in Multiplayer Online Games. In *2015 International Workshop on Network and Systems Support for Games (NetGames)*. IEEE, Zagreb, Croatia, 1–6. <https://doi.org/10.1109/NetGames.2015.7382991> Best Paper Award.
- [53] Ariadna Matamoros-Fernández and Johan Farkas. 2021. Racism, hate speech, and social media: A systematic review and critique. *Television & new media* 22, 2 (2021), 205–224.
- [54] 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. <https://doi.org/10.1145/3275116.3275149>
- [55] Dave McLean, Frank Waddell, and James Ivory. 2020. Toxic Teammates or Obscene Opponents? Influences of Cooperation and Competition on Hostility between Teammates and Opponents in an Online Game. *Journal of Virtual Worlds Research* 13, 1 (2020), 1–15. <https://doi.org/10.4101/jvwr.v13i1.7334>
- [56] Philipe Melo and Clara Pimentel. 2022. A Campanha de Ódio contra The Last of Us Part II. In *Anais Estendidos do XXI Simpósio Brasileiro de Jogos e Entretenimento Digital*. SBC, Natal, RN, Brasil, 428–437. https://doi.org/10.5753/sbgames_estendido.2022.226031
- [57] Luke Munn. 2023. Toxic Play: Examining the Issue of Hate within Gaming. *First Monday* 28, 9 (2023). <https://doi.org/10.5210/fm.v28i9.12508>
- [58] Joaquim A. M. Neto, Kazuki M. Yokoyama, and Karin Becker. 2017. Studying Toxic Behavior Influence and Player Chat in an Online Video Game. In *Proceedings of the 2017 IEEE/WIC/ACM International Conference on Web Intelligence* (Leipzig, Germany) (WI '17). Association for Computing Machinery, New York, NY, USA, 26–33. <https://doi.org/10.1145/3106426.3106452>
- [59] NewZoo. 2017. Male and Female Gamers: How Their Similarities and Differences Shape the Games Market. Newzoo. <https://newzoo.com/resources/blog/male-and-female-gamers-how-their-similarities-and-differences-shape-the-games-market>
- [60] NewZoo. 2024. *2024 Global Games Market Report | Free Version*. Technical Report. Newzoo. <https://newzoo.com/resources/trend-reports/newzoos-global-games-market-report-2024-free-version> Acessado em: 08/08/2025.
- [61] Minh Hao Nguyen, Jonathan Gruber, Will Marler, Amanda Hunsaker, Jaelle Fuchs, and Eszter Hargittai. 2022. Staying Connected While Physically Apart: Digital Communication When Face-to-Face Interactions Are Limited. *New Media & Society* 24, 9 (2022), 2046–2067. <https://doi.org/10.1177/1461444820985442>
- [62] João Gabriel Nogueira. 2023. Review bombing: o que é, como reconhecer e... Tem jeito de impedir? <https://meups.com.br/especiais/review-bombing-o-que-reconhecer-impedir>
- [63] League of Legends Wiki. 2020. The Tribunal. https://leagueoflegends.fandom.com/wiki/The_Tribunal
- [64] Stephanie M. Ortiz. 2019. "You Can Say I Got Desensitized to It": How Men of Color Cope with Everyday Racism in Online Gaming. *Sociological Perspectives* 62, 4 (2019), 572–588. <https://doi.org/10.1177/073121419837588>
- [65] Juan Carlos Pereira-Kohatsu, Lara Quijano-Sánchez, Federico Liberatore, and Miguel Camacho-Collados. 2019. Detecting and Monitoring Hate Speech in Twitter. *Sensors* 19, 21 (2019), 4654. <https://doi.org/10.3390/s19214654>
- [66] Clara Andrade Pimentel and Philipe de Freitas Melo. 2020. Como o Game Design Pode Incentivar o Comportamento Tóxico em Jogos Online. In *Anais do XIX SBGames – Simpósio Brasileiro de Jogos e Entretenimento Digital*. Sociedade Brasileira de Computação (SBC), Recife, PE, Brasil, 448–457. <https://www.sbgames.org/proceedings2020/CulturaFull/209538.pdf> Trilha Cultura – Full Papers.
- [67] Gshow Podcasts. 2021. Você sabe o que é 'Review Bombing'? Entenda a 'explosão de críticas' que afeta filmes como 'Eternos' e 'Marighella'. <https://gshow.globo.com/podcast/cena-aberta/noticia/voce-sabe-o-que-e-review-bombing-entenda-a-explosao-de-criticas-que-afeta-filmes-como-eternos-e-marighella.ghml>
- [68] Susanne Poeller, Martin Johannes Dechant, Madison Klarkowski, and Regan L. Mandryk. 2023. Suspecting Sarcasm: How League of Legends Players Dismiss Positive Communication in Toxic Environments. *Proceedings of the ACM on Human-Computer Interaction* 7, CHI PLAY, Article 374 (2023), 26 pages. <https://doi.org/10.1145/3611020>
- [69] Elizabeth Reid, Regan L. Mandryk, Nicole A. Beres, Madison Klarkowski, and J. Frommel. 2022. 'Bad Vibrations': Sensing Toxicity From In-Game Audio Features. *IEEE Transactions on Games* 14, 4 (2022), 558–568. <https://doi.org/10.1109/TG.2022.3176849>
- [70] Elizabeth Reid, Regan L. Mandryk, Nicole A. Beres, Madison Klarkowski, and Julian Frommel. 2022. Feeling Good and In Control: In-game Tools to Support Targets of Toxicity. *Proceedings of the ACM on Human-Computer Interaction* 6, CHI PLAY, Article 235 (2022), 27 pages. <https://doi.org/10.1145/3549498>
- [71] Manoel Horta Ribeiro, Pedro H. Calais, Yuri A. Santos, Virgilio A. F. Almeida, and Wagner Meira Jr. 2018. Characterizing and Detecting Hateful Users on Twitter. In *Proceedings of the Twelfth International AAAI Conference on Web and Social Media (ICWSM 2018)* (Stanford, CA, USA). AAAI Press, Palo Alto, CA, USA, 676–679. <https://ojs.aaai.org/index.php/ICWSM/article/view/15057>
- [72] 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. <https://doi.org/10.5753/webmedia.2024.241744>
- [73] Victoria L. Rubin and Sarah C. Camm. 2013. Deception in Video Games: Examining Varieties of Griefing. *Online Information Review* 37, 3 (2013), 369–387. <https://doi.org/10.1108/OIR-10-2011-0181>
- [74] Bárbara Sacchitiello. 2018. O que Pablo, vestidos inteligentes e games têm em comum? <https://cannes.meioemensagem.com.br/noticias-2018/2018/06/13/o-que-pablo-vestidos-inteligentes-e-games-tem-em-comum>
- [75] Kelsea Schulenberg, Gua Freeman, Lingyuan Li, and Catherine Barwulor. 2023. "Creepy Towards My Avatar Body, Creepy Towards My Body": How Women Experience and Manage Harassment Risks in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2, Article 236 (2023), 29 pages. <https://doi.org/10.1145/3610027>
- [76] Sercan Sengün, Joni Salminen, Peter Mawhorter, Soon-gyo Jung, and Bernard J. Jansen. 2019. Exploring the Relationship between Game Content and Culture-Based Toxicity. In *Proceedings of the 30th ACM Conference on Hypertext and Social Media* (Hof, Germany) (HT '19). Association for Computing Machinery, New York, NY, USA, 87–95. <https://doi.org/10.1145/334220.3343652>
- [77] Cuihua Shen, Qiusi Sun, Taeyoung Kim, Grace Wolff, Rabindra Ratan, and Dmitri Williams. 2020. Viral Vitriol: Predictors and Contagion of Online Toxicity in World of Tanks. *Computers in Human Behavior* 108 (2020), 106343. <https://doi.org/10.1016/j.chb.2020.106343>
- [78] Miia Siitila and Ellinoora Havaste. 2018. "A pure meritocracy blind to identity": Exploring the Online Responses to All-Female Teams in Reddit. In *Proceedings of DiGRA 2018 Conference: The Game is the Message*. Digital Games Research Association (DiGRA), Turin, Italy, 17 pages. <https://dl.digra.org/index.php/dl/article/view/947>
- [79] SX Group and Go Gamers. 2024. *PGB2024 Pesquisa Game Brasil / Edição Grátis 2024*. Technical Report. Sioux Group and Go Gamers. <https://www.pesquisagamebrasil.com.br/gratis/painel-gratis-2024> Acessado em: 08/08/2025.
- [80] T. L. Taylor. 2006. *Play Between Worlds: Exploring Online Game Culture*. The MIT Press, Cambridge, MA, USA. <https://doi.org/10.7551/mitpress/5418.001.0001>
- [81] Andrea C. Tricco, Erin Lillie, Wasifa Zarin, Kelly K. O'Brien, Heather Colquhoun, Danielle Levac, David Moher, Micah D.J. Peters, Tanya Horsley, Laura Weeks, Susanne Hempel, Elie A. Akl, Christine Chang, Jessie McGowan, Lesley Stewart, Lisa Hartling, Adrian Aldcroft, Michael G. Wilson, Chantelle Garrity, Simon Lewin, Christina M. Godfrey, Marilyn T. Macdonald, Etienne V. Langlois, Karla Soares-Weiser, Jo Moriarty, Tammy Clifford, Özge Tunçalp, and Sharon E. Straus. 2018. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine* 169, 7 (2018), 467–473. <https://doi.org/10.7307/9781536800820-007>

- 10.7326/m18-0850
- [82] Louise H. Trudgett-Klose and Sarven S. McLinton. 2024. "Pro Gamers" & Cyberbullying: Workplace Bullying Sexual Harassment in Professional Video Gaming. *Entertainment Computing* 50 (2024), 100702. <https://doi.org/10.1016/j.entcom.2024.100702>
- [83] Selen Türkay, Jessica Formosa, Sonam Adinolfi, Robert Cuthbert, and Roger Altizer. 2020. See No Evil, Hear No Evil, Speak No Evil: How Collegiate Players Define, Experience and Cope with Toxicity. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA (virtual)) (*CHI '20*). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376191>
- [84] Josiane Van Dorpe, Zachary Yang, Nicolas Grenon-Godbout, and Grégoire Winterstein. 2023. Unveiling Identity Biases in Toxicity Detection: A Game-Focused Dataset and Reactivity Analysis Approach. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track*, Mingxuan Wang and Imed Zitouni (Eds.). Association for Computational Linguistics, Singapore, 263–274. <https://doi.org/10.18653/v1/2023.emnlp-industry.26>
- [85] Maria Drosila Vasconcelos. 2002. Pierre Bourdieu: A herança sociológica. *Educação & Sociedade* 23, 78 (2002), 77–87. <https://doi.org/10.1590/S0101-73302002000200006>
- [86] Michel Wijkstra, Katja Rogers, Regan L. Mandryk, Remco C. Veltkamp, and Julian Frommel. 2024. How To Tame a Toxic Player? A Systematic Literature Review on Intervention Systems for Toxic Behaviors in Online Video Games. *Proceedings of the ACM on Human-Computer Interaction* 8, CHI PLAY, Article 315 (2024), 32 pages. <https://doi.org/10.1145/3677080>
- [87] Tom Wijman. 2018. Mobile Revenues Account for More Than 50% of the Global Games Market as It Reaches \$137.9 Billion in 2018. Newzoo Blog. <https://newzoo.com/resources/blog/global-games-market-reaches-137-9-billion-in-2018-mobile-games-take-half> Published Apr 30, 2018.