



INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO PIAUÍ
CAMPUS ANGICAL
CURSO DE TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

JOÃO GABRIEL NEPOMUCENO DE SOUSA

USOS DA INTELIGÊNCIA ARTIFICIAL NA VIOLÊNCIA DE GÊNERO DIGITAL:
produção, amplificação e enfrentamento

ANGICAL DO PIAUÍ
2026

JOÃO GABRIEL NEPOMUCENO DE SOUSA

USOS DA INTELIGÊNCIA ARTIFICIAL NA VIOLÊNCIA DE GÊNERO DIGITAL:
produção, amplificação e enfrentamento

Trabalho de Conclusão de Curso (artigo científico) apresentado como exigência parcial para obtenção do diploma do Curso de Tecnologia em Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Campus Angical.

Orientador: Prof. Me. José Soares Da Silva Neto

USOS DA INTELIGÊNCIA ARTIFICIAL NA VIOLÊNCIA DE GÊNERO DIGITAL: produção, amplificação e enfrentamento

João Gabriel Nepomuceno de Sousa¹

Prof. Me. José Soares da Silva Neto²

RESUMO

Este trabalho analisa como a inteligência artificial tem sido utilizada na produção, amplificação e no enfrentamento da violência de gênero contra a mulher no ambiente digital. A pesquisa possui abordagem qualitativa, exploratória e descritiva, desenvolvida por meio de pesquisa bibliográfica e documental. Foram realizadas buscas nas bases SciELO, Springer e Google Acadêmico, que resultaram em 57 artigos e documentos, dos quais 17 foram selecionados para análise conforme os critérios de inclusão e exclusão estabelecidos. Os resultados demonstram que a inteligência artificial não criou a violência de gênero digital, mas aumentou sua capacidade de produção e circulação. Ferramentas generativas facilitam a criação de *deepfakes* sexuais, imagens de “nudificação” e outros conteúdos falsos ou manipulados, enquanto sistemas automatizados e algoritmos de recomendação podem ampliar o alcance e a repetição das agressões. Em contrapartida, técnicas de processamento de linguagem natural, aprendizado de máquina e visão computacional podem auxiliar na detecção de misoginia, assédio, discurso de ódio e conteúdos manipulados, além de apoiar a moderação e a orientação das vítimas. Contudo, essas aplicações apresentam limitações relacionadas a vieses, interpretação do contexto e erros de classificação. Conclui-se que o enfrentamento do fenômeno requer, além de recursos tecnológicos com supervisão humana, legislação adequada, responsabilização das plataformas, prevenção, canais de denúncia, apoio às vítimas e respeito ao consentimento.

Palavras-chave: inteligência artificial; violência de gênero digital; *deepfake*; mídias digitais.

ABSTRACT

This study analyzes how artificial intelligence has been used in the production, amplification, and combating of digital gender-based violence against women. The research adopts a qualitative, exploratory, and descriptive approach, based on bibliographic and documentary research. Searches were conducted in the SciELO, Springer, and Google Scholar databases, resulting in 57 articles and documents, of which 17 were selected for analysis according to the established inclusion and exclusion criteria. The findings demonstrate that artificial intelligence did not create digital gender-based violence but increased its capacity for production and dissemination. Generative tools facilitate the creation of sexual deepfakes, “nudified” images, and other false or manipulated content, while automated systems and recommendation

¹ Graduando em Tecnologia em Análise e Desenvolvimento de Sistemas. Instituto Federal do Piauí (IFPI). E-mail: srnep8@gmail.com

² Mestre em Computação (UFC). Professor no Instituto Federal do Piauí (IFPI). E-mail: jose.neto@ifpi.edu.br

algorithms can expand the reach and repetition of abuse. Conversely, natural language processing, machine learning, and computer vision techniques can assist in detecting misogyny, harassment, hate speech, and manipulated content, as well as supporting content moderation and victim guidance. However, these applications have limitations related to bias, contextual interpretation, and classification errors. It is concluded that combating this phenomenon requires not only technological resources combined with human oversight, but also appropriate legislation, platform accountability, preventive measures, reporting channels, victim support, and respect for consent.

Keywords: artificial intelligence; digital gender-based violence; deepfakes; digital media.

Data de aprovação: 29/07/2026

1 INTRODUÇÃO

No que diz respeito aos contextos históricos, culturais e sociais que estruturam as formas de convívio humano, a relação de gênero assume papel central na organização das relações sociais. Na sociedade contemporânea, essa relação está diretamente associada às dinâmicas de desigualdade que se manifestam na violência de gênero. Esse marcador hierárquico estabelece relações desiguais entre pessoas de diferentes gêneros, tendo a mulher, historicamente, como o principal indivíduo desfavorecido e alvo de ataques políticos, sociais e econômicos. Conforme discute Bandeira (2014), a relação de gênero nasce de estruturas hierárquicas baseadas na condição de sexo. Dessa forma, a violência de gênero contra a mulher configura-se como um fenômeno histórico profundamente enraizado nas estruturas sociais.

O Art. 7º da Lei Maria da Penha (Brasil, 2006) define as seguintes formas de violência de gênero contra a mulher: física, caracterizada por agressões corporais; psicológica, que provoca danos emocionais por meio de ameaças, humilhações ou constrangimentos; patrimonial, relacionada à destruição, retenção ou subtração de bens; sexual, envolvendo qualquer ato sexual sem consentimento; e moral, que inclui calúnia, difamação ou injúria. Essas formas de violência evidenciam como a desigualdade de gênero se manifesta estruturalmente na sociedade, atingindo a mulher em razão de sua condição de gênero. Embora essa legislação tenha sido concebida a partir de contextos presenciais, o avanço das tecnologias digitais e da comunicação modificou os padrões de interação social, permitindo que essas relações e práticas também se reproduzam nos ambientes virtuais.

No contexto das mídias digitais, as práticas de violência e desrespeito às mulheres não se diferenciam substancialmente daquelas observadas no mundo físico, manifestando-se por meio de assédio, perseguição, ameaças, difamação e disseminação de discursos misóginos. Entretanto, nesse ambiente, essas práticas passam a ser potencializadas por recursos tecnológicos que ampliam sua escala, velocidade e alcance. Nesse cenário, o uso da inteligência artificial tem se destacado tanto como um fator de agravamento dessas violências, ao possibilitar

a automação e a amplificação de ataques, quanto como um elemento central nos debates sobre a mediação e o controle dessas práticas nas plataformas digitais.

A inteligência artificial, a princípio, foi desenvolvida como um campo científico voltado ao apoio ao ser humano na resolução de problemas lógico-matemáticos, na automação de tarefas intelectuais repetitivas, como o processamento de linguagem e a análise de dados, bem como na tentativa de compreender e simular aspectos da cognição humana. Contudo, no contexto da violência de gênero no ambiente digital, essas tecnologias têm sido utilizadas de maneira problemática, contribuindo para o aumento do volume, da velocidade e da disseminação de ataques nas mídias digitais, o que evidencia um uso ambíguo dessas ferramentas.

Diante desse cenário de crescente violência contra a mulher nas plataformas digitais, associado ao uso de tecnologias baseadas em inteligência artificial e à própria dualidade que marca sua concepção e aplicação, esta pesquisa busca responder à seguinte questão: de que forma a inteligência artificial tem sido utilizada tanto para produção e amplificação das práticas de violência de gênero quanto para o enfrentamento desse fenômeno no ambiente digital?

As estatísticas relacionadas à violência de gênero contra a mulher apresentam-se historicamente alarmantes, evidenciando a persistência desse problema social em diferentes contextos. A Pesquisa Nacional de Violência contra a Mulher (Brasil, 2025) entrevistou 21.641 brasileiras com 16 anos ou mais em todos os estados e no Distrito Federal mostra que, nos últimos 12 meses, 8,8 milhões de mulheres brasileiras, 10% da população feminina brasileira com 16 anos ou mais, contam que sofreram algum tipo de violência digital. Paralelamente, a popularização de tecnologias baseadas em inteligência artificial introduziu novos desafios, ao possibilitar a automação, a amplificação e a disseminação acelerada de práticas violentas no ambiente digital. Diante desse cenário, torna-se relevante aprofundar a análise sobre os impactos, riscos e potencialidades associados ao uso da inteligência artificial no contexto da violência de gênero contra a mulher.

Esta pesquisa apresenta relevância social ao contribuir para a ampliação da compreensão sobre a violência de gênero contra a mulher no ambiente digital, fomentando reflexões e debates sobre o papel da tecnologia na reprodução e no enfrentamento desse fenômeno. Do ponto de vista tecnológico e acadêmico, o estudo também se justifica ao discutir a aplicação da inteligência artificial em contextos socialmente sensíveis, destacando tanto seus efeitos positivos quanto seus riscos, o que pode subsidiar o desenvolvimento de sistemas mais éticos, responsáveis e comprometidos com a redução das desigualdades.

Este trabalho tem como objetivo geral analisar de que forma a inteligência artificial tem sido utilizada tanto na potencialização quanto no enfrentamento da violência de gênero contra a mulher no ambiente digital, considerando seus impactos sociais e tecnológicos.

Como objetivos específicos, busca-se compreender os principais conceitos relacionados à violência de gênero, às mídias sociais e à inteligência artificial, com base em referenciais teóricos, identificar e discutir os mecanismos de produção e reprodução da violência de gênero nas plataformas digitais mediadas por tecnologias de inteligência artificial e analisar

as estratégias de enfrentamento da violência de gênero no ambiente digital que utilizam a inteligência artificial, destacando seus limites e potencialidades e suas dimensões tecnológicas, jurídicas e institucionais.

2 REFERENCIAL TEÓRICO

O referencial teórico desse estudo está organizado para entender a discussão já existente sobre a violência de gênero contra a mulher, sua manifestação nos ambientes digitais e sua relação com a tecnologia e com a inteligência artificial, apresentando e discutindo os principais conceitos relacionado à pesquisa.

2.1 DA VIOLÊNCIA DE GÊNERO CONTRA A MULHER

A violência contra a mulher é uma das principais formas de violência de gênero. A Organização Pan-Americana da Saúde (OPAS, 2020) classifica a violência contra as mulheres como qualquer ato de violência de gênero que resulte ou possa resultar em danos ou sofrimentos físicos, sexuais ou mentais para as mulheres, em vida pública ou privada. A violência de gênero, por sua vez, ultrapassa as mulheres, produzindo-se e reproduzindo-se nas relações de poder onde entrelaçam as categorias de gênero, classe e raça ou etnia (Araújo, 2008). Portanto, a violência contra a mulher evidencia-se como uma especificidade da violência de gênero recorrente da relação de condição de sexo.

Essa assimetria de poder ocorre por causa de uma hierarquia entre os gêneros do homem e da mulher. Dentre os conceitos e estudos acerca dessa diferença, um que merece destaque é o patriarcado. Os estudos de Bandeira (2014) o considera como um dos pilares para a violência de gênero contra a mulher, com Saffioti (2015) o caracterizando como um regime da dominação-exploração das mulheres pelos homens. Isso significa que existe uma relação desigual dos homens sobre as mulheres naturalizada pela construção social do que é ser homem e do que é ser mulher, fazendo com que o homem se beneficie dessa dominação. Saffioti (2015) já aponta o controle da sexualidade feminina como um dos elementos nucleares do patriarcado. "Ora, a mulher sempre foi, senão a escrava do homem ao menos sua vassala; os dois sexos nunca partilharam o mundo em igualdade de condições", (Beavouir, 1970, p. 14).

2.2 DA VIOLÊNCIA DE GÊNERO CONTRA A MULHER NAS MÍDIAS SOCIAIS

A Lei Maria da Penha (Brasil, 2006), como explicado anteriormente, tem definidas as categorias físico, psicológicas, patrimonial, sexual ou moral como itens de violência de gênero contra a mulher. Por ser pensada em contextos presenciais, essa legislação não faz a proteção em todos os espaços, sendo um desses lugares as plataformas digitais.

Nesse sentido, organismos internacionais passaram a discutir e conceituar especificamente as manifestações de violência de gênero nos ambientes digitais. A Organização dos Estados Americanos (OEA) tem a seguinte definição de violência

digital de gênero contra a mulher:

Qualquer ato, conduta ou omissão contra mulheres, com base no gênero, que cause morte, dano ou sofrimento físico, sexual, psicológico, político ou econômico, incluindo danos patrimoniais, em qualquer esfera de suas vidas, cometido, instigado, facilitado ou agravado, total ou parcialmente, pelo uso de tecnologias digitais. (Organização dos Estados Americanos, 2025).

Esse trecho do documento evidencia que o tema é reconhecido entre entidades e Estados, sendo, assim, uma importante ramificação da violência de gênero online cometida contra a mulher. Contudo, além da definição, é preciso entender de que formas essa violência ocorre nesses espaços.

Sendo assim, Alwis enfatiza os seguintes modos como violência de gênero online:

A violência de gênero online é incorporada de várias maneiras, como perseguição cibernética, assédio cibernético, divulgação não autorizada de imagens íntimas, investigação digital, reprovação de conduta, pegadinhas, exibição cibernética, discurso de ódio misógino, desinformação, informação falsa, campanhas de difamação cibernética, ameaças de violência sexual e de assassinato e manipulação de imagens, como o proliferação de mídia sexualmente explícita gerada por IA. (Alwis, 2024, p. 63, tradução nossa).

As formas de agressões citadas, como o assédio, divulgação de imagens íntimas e outras, realçam a caracterização dessa forma de violência nas mídias sociais, mostrando como isso ocorre no cotidiano digital. Essa potencialização da violência de gênero trazida pelas mídias sociais acontece por causa de uma série de fatores. A União Europeia (2024a) aponta, entre os principais fatores, a difusão fácil, rápida e ampla do discurso de ódio através da palavra digital, reforçada pelo efeito de desinibição em linha, uma vez que o presumível anonimato na internet e o sentimento de impunidade reduzem a inibição das pessoas para participar nesse tipo de discurso.

Alwis (2024) também discorre sobre isso, enfatizando que o anonimato dos criminosos em relação às vítimas, a distância geográfica e a capacidade de automação oferecida pela tecnologia amplifica o escopo e o impacto do comportamento abusivo. Nesse sentido, a internet e as redes digitais evidenciam-se como agentes potencializadores da violência de gênero contra a mulher, levando-a para outros tipos de espaços.

Como resposta, o Estado tem tentado intervir com leis. Uma lei popularmente conhecida como Lei Carolina Dieckmann (Brasil, 2012), que prevê detenção para invasões a dispositivos informáticos alheios sem devida autorização, foi criada após a atriz Carolina Dieckmann ter sua intimidade violada quando um grupo de hackers invadiu seu computador pessoal e divulgaram, sem autorização, 36 imagens íntimas pelas redes sociais (Defensoria Pública do Estado do Ceará, 2022).

Também existe a Lei nº 13.718 (Brasil, 2018), que tipifica o crime de importunação sexual e divulgação de cena de estupro, comunicando-se diretamente com os tipos de violência

de gênero que ocorrem nas mídias sociais. "A possibilidade de punir aqueles que divulgam imagens íntimas sem consentimento, por exemplo, é uma forma de coibir essa prática tão comum na sociedade atual", analisa Amaral (2024, p. 4).

Assim sendo, por mais que o Estado brasileiro e outros órgãos reconheçam a violência baseada em gênero recorrente nas plataformas sociais digitais, ainda há desafios que precisam ser enfrentados e superados.

2.3 DA TECNOLOGIA E INTELIGÊNCIA ARTIFICIAL

Conforme apresentado, a violência de gênero digital é um processo cotidiano na vida de quem usa as redes sociais, mas o fato dela ocorrer não nega o também fato das próprias plataformas digitais tentarem diminuir ou barrar essa violência.

Primeiramente, muitas plataformas já têm seus conjuntos de normas e linhas de condutas que um usuário deve seguir ao usar sua rede. "Esses documentos geralmente contêm as principais diretrizes sobre conteúdo e ações que são permitidas ou não na rede social, tais como violência, assédio, discurso de ódio, notícias falsas, nudez e terrorismo" (Ferraz *et al.*, 2024, p. 385). Isso significa que, quando as diretrizes são violadas, os usuários podem denunciar as publicações digitais ou a própria empresa reagir.

Entretanto, quando esse tipo de conteúdo é denunciado para a empresa ou plataforma, naturalmente vai se tornando um grande volume de denúncias para vigiar, resultando em uma inviabilização do trabalho humano para a moderação desses conteúdos.

O número de conteúdo gerado pelos usuários está crescendo rapidamente, impulsionado pelo fácil acesso a celulares e conexões de internet. Pela mesma razão, o acesso a esse conteúdo está mais fácil do que antes. Usar humanos para moderar esse conteúdo se apresenta como um desafio significativo, dado o volume bruto de conteúdo gerado. (OFCOM, 2019, tradução nossa).

No que se refere ao objeto acusado como denúncia, muitas vezes, elas aparecem em diferentes formatos que requerem análises de contexto ou significados, como vídeos e memes (OFCOM, 2019), ou, até mesmo, os conteúdos que se enquadram como violência de gênero contra a mulher, que é o foco deste estudo. Portanto, em decorrência do volume e dos obstáculos na moderação, a Inteligência Artificial (IA) se torna um importante aliado para o enfrentamento da violência de diretrizes das plataformas sociais na internet.

A IA "é uma tecnologia que permite que computadores e máquinas simulem o aprendizado, a compreensão, a resolução de problemas, a tomada de decisões, a criatividade e a autonomia dos seres humanos" (Stryker; Kavlakoglu, 2024). "É um conjunto de tecnologias que permite que os computadores aprendam, raciocinem e realizem várias tarefas avançadas de maneiras que antes exigiam inteligência humana" (Google, [s.d.]).

Russell e Norvig (2013) articulam que a IA abrange uma enorme variedade de subcampos, do geral (aprendizagem e percepção) até tarefas específicas, como jogos de xadrez,

demonstração de teoremas matemáticos, criação de poesia, direção de um carro em estrada movimentada e diagnóstico de doenças. Também abrange a eliminação de tarefas repetitivas e na aceleração e desenvolvimento de pesquisas ou em áreas, como em operações comerciais, saúde ou no dia a dia do ser humano (Google, [s.d]). Sua independência e raciocínio rápido potencializam todos esses setores.

Graças, também, a essa série de aplicabilidades, não demorou muito para as plataformas usarem a IA para a moderação automatizada dos espaços digitais. Nesse cenário, um dos usos mais comuns é contra o *spam* (Russell; Norvig, 2013), poupando o destinatário e os moderadores das redes sociais de perder tempo excluindo cerca de 80 a 90 por cento das mensagens recebidas.

Essa leitura automatizada de spams só é possível por meio do que é chamado de aprendizado de máquina, que é um processo que envolve treinar uma máquina para fazer previsões ou decisões baseadas em dados (Rashid; Kausik, 2024). Dentre os campos do aprendizado de máquina, uma das abordagens adotadas é o aprendizado supervisionado, que tem como base a rotulação prévia dos dados. Ou seja, nesse contexto, as mensagens são apresentadas com uma indicação de que aquele conteúdo é ou não é spam. Conforme definido pelo Google ([s.d.]), trata-se de um tipo de inteligência artificial no qual os sistemas aprendem com os dados para identificar padrões e realizar previsões ou decisões sem programação direta. Dessa forma, o processo é caracterizado como classificação supervisionada, no qual a IA aprende a categorizar as mensagens a partir de exemplos previamente rotulados. Esse princípio da classificação supervisionada pode ser estendido para outros conteúdos presentes nas plataformas digitais, tais quais as situações descritas como violência de gênero digital.

Entretanto, atualmente, tendo em vista que existem outras formas de violência de gênero que não usam texto para atacar as vítimas, nascem as chamadas *deepfakes*. Elas utilizam IA e tecnologia de mapeamento facial para mesclar, combinar e sobrepor imagens e vídeos. Sob esse contexto, essa criação de mídias falsas pode gerar o chamado de *deepfakes* pornográficos, reforçando uma cultura de mercantilização e objetificação do corpo das mulheres (Alwis, 2024) resultando na humilhação e violência digital contra a mulher.

Essa criação de vídeos e imagens falsas só é possível graças ao que é chamado de *deep learning* (aprendizado profundo) e IA generativa. O aprendizado profundo é um outro subcampo do aprendizado de máquina, que usa redes neurais multicamadas, chamadas de redes neurais profundas, que simulam de forma mais aproximada o complexo poder de tomada de decisões do cérebro humano (Stryker; Kavlakoglu, 2024). Essa subcamada da IA é utilizada pela IA generativa, que, por sua vez, refere-se a computadores capazes de criar conteúdos originais complexos, como textos longos, imagens, vídeos, áudios realistas e muito mais (Stryker; Kavlakoglu, 2024).

Em termos gerais, observa-se que a inteligência artificial não pode ser compreendida como uma tecnologia intrinsecamente benéfica ou prejudicial, mas como uma ferramenta cujo impacto depende dos contextos de uso. Sua ambiguidade permite que seus sistemas

sejam utilizados tanto para identificar e classificar conteúdos prejudiciais nas plataformas digitais quanto para a criação e amplificação desses mesmos conteúdos, evidenciando riscos de humilhação, difamação e violência nos casos de deepfakes e de manipulação de imagens.

3 METODOLOGIA

A metodologia adotada neste trabalho é de natureza qualitativa, com caráter exploratório e descritivo, sendo adequada ao nível do curso de Tecnologia em Análise e Desenvolvimento de Sistemas (TADS).

O estudo foi desenvolvido por meio de pesquisa bibliográfica e documental, com o objetivo de analisar como a IA tem sido empregada tanto na potencialização quanto no enfrentamento da violência de gênero contra a mulher no ambiente digital. A pesquisa bibliográfica consistiu no levantamento, seleção e análise de artigos científicos, livros, dissertações, teses e publicações acadêmicas que abordem os temas de violência de gênero, mídias digitais e IA, priorizando produções recentes e relevantes para a área.

Paralelamente, houve uma pesquisa documental conduzida de forma incremental, à medida que o trabalho demandava instrumentos normativos, sem protocolo prévio equivalente ao da busca bibliográfica, isso à partir da análise de documentos institucionais, legislações, relatórios técnicos, diretrizes de organismos nacionais e internacionais e materiais produzidos por órgãos ligados à segurança pública e à proteção dos direitos das mulheres. Esses documentos foram utilizados para compreender o enquadramento legal, social e tecnológico da violência de gênero no ambiente digital, bem como as estratégias de enfrentamento já existentes.

A aplicação da metodologia e a coleta de dados foram realizadas em três diferentes bases de dados: SciELO, Springer e Google Acadêmico. Dentro dessas bases, foi utilizada uma *string* de busca para pesquisar sobre as bibliografias e documentos necessários na realização desta pesquisa. A montagem dessa *string* foi pensada para que ela traga os documentos necessários que abordem sobre IA, violência de gênero ou violência contra a mulher e violência digital, essa ainda ligada com a IA.

Na SciELO, foi realizada a busca (*artificial intelligence*) AND ((*gender based violence*) OR (*digital violence*)), com os filtros de linguagem em inglês e espanhol e com os anos de publicação de 2022 a 2026, gerando 8 resultados.

No Springer, foi definida a *string* de busca ("*gender based violence*" OR "*violence against women*" OR "*online violence*") AND ("*artificial intelligence*") AND ("*social media*"). Dentro dos filtros disponíveis, foram selecionados artigos, artigos de pesquisa e artigos de revisão, com modelo de publicação de acesso público e com data de publicação de 2022 a 2026. Além disso, foram selecionados os assuntos sobre IA, gênero e tecnologia da informação (TIC), ética digital, gênero e mídia, violência sexual e doméstica, aprendizado de máquina, ética computacional, mineração de dados, dinâmicas e intervenções em casos de violência doméstica, crime computacional e dinâmica dos *deepfakes* na desinformação em mídias sociais,

todos tangendo pelo menos uma área da questão da pesquisa, providenciando 32 resultados.

Por sua vez, foi utilizada a *string* de busca *allintitle: "gender based violence" "artificial intelligence"* no Google Acadêmico, onde as palavras aparecem somente no título do artigo e publicados desde 2022, produzindo 17 resultados.

A análise dos materiais selecionados ocorreu de forma mista, buscando identificar padrões, abordagens recorrentes, lacunas na literatura e diferentes formas de utilização da IA no contexto da violência de gênero digital. Os resultados dessa análise serviram de base para a discussão sobre a dualidade do uso da IA, evidenciando tanto seus riscos quanto suas potencialidades no enfrentamento desse fenômeno.

3.1 PROCEDIMENTOS DA PESQUISA BIBLIOGRÁFICA

Nesta etapa, foram definidos os critérios utilizados para selecionar os estudos que compuseram a pesquisa bibliográfica. O Quadro 1 apresenta os critérios de inclusão e exclusão adotados para a seleção dos trabalhos.

Quadro 1 – Critérios de inclusão e exclusão da pesquisa bibliográfica

Categoria	Descrição
Critérios de Inclusão	Publicações científicas, tais como artigos e revisões sistemáticas; Trabalhos publicados nos últimos cinco anos (2022 - 2026); Estudos que abordem a violência de gênero contra a mulher no ambiente digital; Pesquisas que discutam o uso da IA, de forma direta ou indireta, na produção, amplificação, detecção ou enfrentamento da violência de gênero online; Publicações disponíveis em português, espanhol ou inglês; Textos com acesso ao conteúdo completo.
Critérios de Exclusão	Trabalhos fora do recorte temporal definido (publicações com mais de cinco anos); Estudos que abordem a violência de gênero apenas em contextos presenciais, sem relação com o ambiente digital; Pesquisas que tratem de IA sem vínculo com violência de gênero ou com questões sociais associadas, jornalísticas ou materiais sem caráter científico; Publicações duplicadas em mais de uma base de dados.

Fonte: Elaborado pelo autor (2026).

3.2 PROCEDIMENTOS DA PESQUISA DOCUMENTAL

A pesquisa documental foi realizada a partir da análise de diferentes categorias de documentos relacionados ao tema investigado. O Quadro 2 apresenta as categorias de documentos consideradas e sua respectiva descrição.

Quadro 2 - Categorias de documentos analisados na pesquisa documental

Categoria de Documento	Descrição
Documentos legais e normativos	Leis que tratam da violência de gênero e de crimes digitais, como a Lei Maria da Penha; Diretrizes nacionais e internacionais relacionadas à violência digital e às questões de gênero.
Relatórios institucionais	Relatórios produzidos por órgãos de segurança pública; Documentos e relatórios elaborados por organismos internacionais, como ONU, OEA e União Europeia.
Documentos técnicos de plataformas digitais	Políticas de uso, termos de serviço e diretrizes de moderação de conteúdo das plataformas digitais; Relatórios de transparência sobre o uso de IA e processos de moderação automática.
Documentos técnicos e estratégicos	Guias, notas técnicas e recomendações voltadas ao uso da IA para a segurança no ambiente digital.

Fonte: Elaborado pelo autor (2026).

A soma dos resultados obtidos totaliza 57 artigos e documentos. Desse conjunto, foram excluídos aqueles que atendem aos critérios de exclusão estabelecidos no Quadro 1, bem como os estudos que, após uma leitura prévia do título e do resumo, não apresentam afinidade à questão central desta investigação. Além das buscas nas bases, foi incluído 1 estudo identificado por rastreamento de referências, gerando 17 estudos totais para a análise.

4 RESULTADOS E DISCUSSÃO

Entre os 17 estudos selecionados, foram identificadas revisões bibliográficas, análises jurídicas, estudos qualitativos e pesquisas experimentais voltadas ao desenvolvimento de modelos computacionais. Após a seleção, os estudos foram lidos integralmente e organizados em uma matriz de síntese. Foram extraídas informações referentes à autoria, fonte de recuperação, tipo de estudo, aplicação da inteligência artificial e categoria de enquadramento adotada nesta pesquisa. Posteriormente, os achados foram comparados e agrupados nas categorias compreensão do fenômeno, produção, amplificação, enfrentamento e proteção, além dos aspectos jurídicos, institucionais e sociais. Os trabalhos não apresentaram uma distribuição uniforme entre as categorias: parte considerável concentrou-se na produção de conteúdos sintéticos e nos aspectos jurídicos, enquanto foram encontrados menos estudos dedicados à avaliação prática de ferramentas de proteção às vítimas. Um mesmo estudo pôde integrar mais de uma categoria, conforme suas contribuições. A análise buscou identificar convergências, divergências, limitações e lacunas presentes na literatura.

O quadro abaixo apresenta uma visão geral dos 17 estudos selecionados para compor esta revisão integrativa, sintetizando a fonte de recuperação, o tipo de estudo, a principal aplicação da inteligência artificial investigada e a categoria de enquadramento utilizada na análise.

Quadro 3 – Síntese dos estudos incluídos na revisão integrativa

Autor	Fonte	Tipo de estudo	Aplicação de IA examinada	Categoria(s) de enquadramento
Polyzoidou	Springer	Estudo jurídico	Violência digital contra mulheres, abuso de imagens íntimas e desafios regulatórios envolvendo IA.	Estudos; governança.
Rachmaria e Susanto	Google Acadêmico - rastreamento de referência	Revisão de literatura	Deepfakes pornográficos e proteção jurídica às vítimas.	Estudos.
Bellomo	Google Acadêmico	Revisão teórica	Discriminação algorítmica, pornografia sintética e impactos sociais da IA.	Estudos; produção; amplificação; enfrentamento; governança.
Balavadze	Google Acadêmico	Revisão bibliográfica	IA e violência de gênero, desafios éticos e regulatórios.	Estudos.
Döring <i>et al.</i>	Springer	Revisão de literatura	IA generativa aplicada à sexualidade e pornografia sintética.	Estudos.
Döring, Le e Miller	Springer	Análise quantitativa de conteúdo	Produção e compartilhamento de pornografia gerada por IA em comunidades on-line.	Produção.
Maras e Logie	Springer	Revisão crítica	Deepfakes sexualmente explícitos, danos às vítimas e estratégias de enfrentamento.	Produção; enfrentamento; Estudos; governança.
Kopecký e Voráč	Springer	Revisão de literatura	Deep nudes e pornografia sintética produzida por IA.	Enfrentamento; Estudos.
Diel <i>et al.</i>	Springer	Revisão sistemática	Consequências psicológicas dos deepfakes de conteúdo íntimo.	Estudos; governança.
Dutta, Banducci e Camargo	Springer	Revisão sistemática	Técnicas de IA para detecção automática de sexismo e misoginia online.	Enfrentamento; Estudos.
Imran e Tariq	Google Acadêmico	Revisão narrativa	Utilizações benéficas e maliciosas da IA na violência de gênero.	Produção; amplificação; enfrentamento.
Hashmi <i>et al.</i>	Springer	Estudo experimental	Modelo de aprendizado de máquina para identificação automática de misoginia.	Enfrentamento.

Neubauer <i>et al.</i>	Springer	Revisão sistemática	Métodos computacionais e IA aplicados ao estudo da violência por parceiro íntimo.	Enfrentamento.
Ganguly, Priya e Pandey	Google Acadêmico	Estudo analítico	Aplicações da IA no enfrentamento da violência de gênero facilitada por tecnologias.	Enfrentamento.
Hoegg	Google Acadêmico	Revisão teórica	IA generativa, desinformação, estereótipos e amplificação da violência.	Amplificação.
Lorca e García- Mingo	SciELO	Estudo qualitativo	Violência sexual digital, plataformas digitais e manipulação de imagens por IA.	Amplificação.
Pawelec	Springer	Estudo qualitativo	Desenvolvimento ético e governança de ferramentas de deepfake.	Governança.

Fonte: Elaborado pelo autor (2026).

4.1 COMPREENSÕES SOBRE IA E VIOLÊNCIA DE GÊNERO DIGITAL

Os estudos analisados convergem na compreensão de que a violência de gênero digital não surgiu com a IA, mas deriva das desigualdades e formas de violência já existentes na sociedade (Polyzoidou, 2024). Essa compreensão também encontra respaldo em Araújo (2008), ao destacar que a violência de gênero decorre das relações historicamente desiguais entre homens e mulheres, sendo atravessada por marcadores sociais como classe e raça. Sob essa perspectiva, as tecnologias digitais e a inteligência artificial não constituem a origem dessas violências, mas reproduzem e potencializam desigualdades estruturais já presentes na sociedade. Essa abordagem interseccional permite compreender por que determinados grupos, como mulheres negras, pessoas LGBTQIAPN+, minorias linguísticas e outros segmentos socialmente marginalizados, tendem a experimentar formas mais intensas e complexas de violência nos ambientes digitais, evidenciando que os impactos da inteligência artificial são condicionados pelas desigualdades sociais preexistentes. Embora as tecnologias digitais não criem essas relações desiguais entre homens e mulheres (Rachmaria; Susanto, 2024), elas oferecem novos recursos para reproduzir práticas de controle, humilhação, perseguição e objetificação (Bellomo, 2025), ampliando o alcance, a reprodução e a permanência dos conteúdos violentos.

A IA também não é neutra, pois seus sistemas são construídos a partir de dados, decisões humanas, interesses empresariais e estruturas sociais marcadas por desigualdades, podendo reproduzir preconceitos, estereótipos e associações discriminatórias (Bellomo, 2025). Contudo, a IA não constitui, por si só, causa ou solução para a violência de gênero, pois seus efeitos

dependem de como é projetada e utilizada (Balavadze, 2025). Os danos, portanto, não decorrem somente da tecnologia, mas também de sua conexão com estruturas sociais que normalizam a objetificação e o controle das mulheres (Rachmaria; Susanto, 2024).

Os conteúdos íntimos falsos produzidos por IA devem ser compreendidos como uma forma de violência baseada em imagens, mesmo quando não representam uma situação real. Döring *et al.* (2025) utilizam a expressão "abuso sexual baseado em imagens geradas por IA" para destacar que o problema central é a utilização sexual da identidade de uma pessoa sem consentimento. Polyzoidou (2024) também considera o consentimento fundamental, independentemente de a imagem ser originalmente íntima, posteriormente manipulada ou inteiramente produzida por meios digitais, pois esse uso viola a dignidade e a autodeterminação sexual da vítima. Maras e Logie (2024) acrescentam que cada pessoa deve controlar o uso de sua imagem e voz no ambiente digital. Da mesma forma, Kopecký e Voráč (2026) apresentam os *deep nudes* como uma nova forma de violência capaz de atingir pessoas de diferentes faixas etárias. Assim, a falsidade do conteúdo não elimina a violência nem os danos causados à reputação da vítima.

Diel *et al.* (2026) reforçam essa compreensão ao identificarem, entre vítimas de abuso sexual baseado em imagens envolvendo *deepfakes*, consequências como medo, vergonha, culpa, isolamento, depressão, insônia e perda de apetite. Embora ainda existam poucos estudos empíricos sobre esses impactos, os resultados disponíveis demonstram danos reais decorrentes da apropriação da identidade das vítimas.

A violência de gênero digital também não se limita a ataques explicitamente ofensivos. Dutta, Banducci e Camargo (2025) mostram que o sexismo e a misoginia podem aparecer de maneira sutil, por meio de ironias, piadas, estereótipos, comentários aparentemente protetores ou discursos que negam a desigualdade de gênero. Segundo os autores, parte das pesquisas computacionais reduz esse fenômeno a classificações como “sexista” ou “não sexista”, desconsiderando seu contexto social e cultural. Por isso, defendem a aproximação entre computação, ciências sociais, linguística e estudos de gênero, pois somente o aprimoramento dos algoritmos não garante a compreensão adequada dos conteúdos analisados.

Em síntese, a relação entre IA e violência de gênero digital constitui um fenômeno social e tecnológico. Seus efeitos resultam da combinação entre as capacidades da tecnologia, os dados utilizados, as escolhas dos desenvolvedores, as ações dos usuários, o funcionamento das plataformas e as desigualdades sociais. Dessa forma, a IA não é apresentada como essencialmente maléfica ou benéfica, mas como uma ferramenta cujos efeitos dependem dos objetivos, valores e mecanismos de controle envolvidos em seu desenvolvimento e utilização.

4.2 USO DA IA NA PRODUÇÃO DA VIOLÊNCIA DE GÊNERO DIGITAL

Dentro da literatura analisada, a IA é utilizada na produção da violência de gênero digital principalmente para criar conteúdos falsos ou manipulados que associam mulheres e meninas a

situações sexuais sem consentimento, tópico já discutido na seção 4.1. Imran e Tariq (2025) apontam que ferramentas de IA generativa produzem imagens, vídeos, áudios e narrativas capazes de representar vítimas realizando ações que nunca aconteceram. Nesse contexto, as *deepfakes* ocupam posição central por utilizarem a identidade de uma pessoa na fabricação de conteúdos abusivos aparentemente reais.

A produção de pornografia artificial envolve diferentes áreas da IA. Döring, Le e Miller (2025) destacam a IA generativa, enquanto Maras e Logie (2024) apontam a participação do aprendizado de máquina. Essas tecnologias permitem gerar conteúdos automaticamente, substituir rostos, modificar corpos e sintetizar registros aparentemente autênticos. Döring, Le e Miller (2025) também identificaram a engenharia de *prompts* como parte desse processo, pois os comandos permitem definir características do corpo, do rosto, do cenário e da situação sexual representada. Entretanto, o tema apareceu em apenas 2,3% das 390 postagens analisadas, indicando que a elaboração desses comandos foi pouco discutida. O estudo ainda menciona *chatbots* generativos capazes de produzir interações sexuais personalizadas por texto ou áudio, com aparência, voz e personalidade configuradas pelo usuário.

Essas aplicações também reduziram as dificuldades técnicas da manipulação de imagens. Bellomo (2025) destaca aplicativos de troca de rostos e “nudificação” que transformam fotografias comuns em representações sexuais falsas com poucos comandos. Assim, imagens publicadas em redes sociais podem servir como matéria-prima para conteúdos íntimos artificiais sem consentimento. A autora relata que um vídeo pornográfico *deepfake* de sessenta segundos poderia ser produzido gratuitamente com apenas uma imagem clara do rosto da vítima. Maras e Logie (2024) também observam que muitas ferramentas são gratuitas ou exigem um pequeno pagamento e poucas imagens para gerar resultados convincentes. Portanto, a IA reduz o tempo, o custo e o conhecimento técnico necessários para essa prática, tornando-a acessível a mais usuários.

Essa aproximação entre usuários comuns e ferramentas de produção também aparece no estudo de Döring, Le e Miller (2025), que analisou 390 postagens públicas sobre pornografia gerada por IA no Reddit, publicadas entre 2016 e 2024. Em 59,5% delas havia discussões sobre a produção desse conteúdo, feitas por produtores, observadores ou vítimas. Entre os temas identificados, 28,5% mencionavam a criação de conteúdos específicos, 23,8% discutiam as ferramentas utilizadas, 15,9% abordavam a produção em geral e 2,3% tratavam da elaboração de *prompts*. Também havia pedidos de recomendações e divulgação de serviços comerciais. Além disso, 39,5% das postagens tratavam especificamente de pornografia *deepfake*, envolvendo não apenas celebridades, mas colegas de trabalho, amigas, vizinhas, ex-parceiras e familiares. Embora esses resultados não comprovem que todas as ferramentas sejam simples ou baratas, demonstram que a produção já ocorre em relações cotidianas e não está restrita a especialistas.

Crianças e adolescentes também são atingidos por essa violência. Bellomo (2025) reúne casos em que estudantes utilizaram aplicativos de “nudificação” para criar imagens sexuais falsas de colegas. A autora também menciona a crise ocorrida na Coreia do Sul, na qual mais

de quinhentas instituições educacionais e aproximadamente dois mil grupos de conversa foram relacionados à produção e à circulação de pornografia *deepfake*, envolvendo principalmente adolescentes. Esses casos mostram que a violência pode ser praticada por pessoas inseridas nos círculos escolar, familiar, afetivo ou profissional das vítimas.

Em síntese, o elemento determinante não é apenas o uso da IA ou a natureza pornográfica do conteúdo, mas a apropriação não consentida da identidade da vítima, sua sexualização e a finalidade abusiva da produção. A principal transformação provocada pela IA está na possibilidade de fabricar conteúdos sexuais, difamatórios ou ofensivos com maior realismo, personalização, rapidez e facilidade, utilizando a identidade e os registros digitais das vítimas como matéria-prima.

4.3 USO DA IA NA AMPLIFICAÇÃO DA VIOLÊNCIA DE GÊNERO DIGITAL

Além de produzir conteúdos abusivos, a IA pode participar da amplificação da violência de gênero digital, entendida neste trabalho como o aumento da quantidade, da frequência, do alcance ou da repetição das agressões. Esse processo pode ocorrer diretamente pela IA, indiretamente pelo aumento do volume de conteúdos disponíveis ou externamente, por meio das plataformas e da atuação coletiva dos usuários.

A amplificação ocorre diretamente quando a IA aumenta a escala ou a frequência da violência. Imran e Tariq (2025) mencionam sistemas automatizados utilizados em campanhas de assédio, capazes de produzir e enviar mensagens abusivas repetidamente. Esses sistemas também podem criar diferentes versões de imagens, áudios, textos e narrativas falsas para humilhar, difamar ou prejudicar a reputação das vítimas. Nesses casos, a IA participa simultaneamente da produção e da amplificação ao multiplicar os ataques e acelerar sua disseminação.

Outro mecanismo direto está nos sistemas algorítmicos de recomendação. Bellomo (2025) explica que esses sistemas priorizam publicações com potencial de gerar interação e engajamento. Ao recomendar conteúdos misóginos, discriminatórios ou sexualmente abusivos, os algoritmos ampliam sua visibilidade e podem levá-los a pessoas que não faziam parte do público inicial.

A reprodução sistemática de preconceitos também constitui uma forma de amplificação. Hoegg (2025) explica que modelos generativos treinados com conteúdos da internet podem aprender discursos de ódio e estereótipos de gênero. Imran e Tariq (2025) acrescentam que bases pouco representativas e equipes com pouca diversidade dificultam a identificação desses vieses. Uma resposta preconceituosa isolada representa a produção de conteúdo discriminatório, enquanto a amplificação ocorre quando esse padrão é repetido automaticamente em numerosas interações. Hoegg (2025) relaciona essa repetição à permanência da violência estrutural, simbólica e "lenta", pois contribui para normalizar desigualdades já existentes.

A segunda forma ocorre quando a IA não dissemina diretamente o conteúdo, mas

aumenta a quantidade de material abusivo disponível. Bellomo (2025) explica que ferramentas generativas reduziram o tempo, o custo e o conhecimento técnico necessários para manipular imagens, vídeos, áudios e textos. Aplicativos de troca de rostos e "nudificação", por exemplo, permitem criar representações sexuais falsas com poucos comandos. Maras e Logie (2024) também observam que muitas dessas ferramentas são gratuitas ou relativamente baratas e precisam de poucas imagens para gerar resultados convincentes. Assim, embora o compartilhamento seja realizado por usuários ou plataformas, a IA favorece sua amplificação ao permitir que mais conteúdos sejam produzidos em menos tempo e por mais pessoas.

A terceira forma reúne situações em que a IA pode produzir parte do material, mas sua amplificação é realizada pelas plataformas e pelos usuários. Lorca e García-Mingo (2023) analisaram a Hispasexy, comunidade digital masculina voltada à obtenção e ao compartilhamento de imagens íntimas de mulheres sem consentimento. Entre os materiais havia conteúdos retirados de redes sociais, arquivos vazados e imagens falsas produzidas com IA. Entretanto, o estudo não afirma que a própria comunidade oferecia ferramentas de IA: a IA aparecia na fabricação de algumas imagens, enquanto sua circulação dependia da plataforma e de seus integrantes.

As funcionalidades da Hispasexy facilitavam a busca, a catalogação, o download e o armazenamento externo dos arquivos, permitindo que continuassem disponíveis mesmo após a exclusão da publicação original. Os usuários também solicitavam novos materiais, incentivavam outros integrantes e mantinham os conteúdos em circulação. Por isso, Lorca e García-Mingo (2023) caracterizam a comunidade como um "arquivo permanente de violência", no qual imagens íntimas se tornavam troféus coletivos e fortaleciam os vínculos entre os agressores.

Os autores denominam essa dinâmica de misoginia colaborativa, caracterizada pela cooperação dos membros na obtenção, fabricação, solicitação, organização, conservação e circulação das imagens. A exposição ainda poderia gerar novas agressões, como humilhação direta das vítimas, perseguição, divulgação de informações pessoais, ameaças e outras formas de assédio. Nesse caso, embora a IA pudesse produzir ou modificar algumas imagens, eram os usuários e as funcionalidades da plataforma que prolongavam sua existência e provocavam a revitimização.

A compreensão da misoginia colaborativa também pode ser na concepção de patriarcado como um regime de dominação-exploração proposta por Saffioti (2015). Nesse sentido, as práticas de violência observadas nos ambientes digitais não representam fenômenos isolados ou decorrentes exclusivamente das tecnologias, mas constituem uma continuidade das relações sociais historicamente estruturadas pela desigualdade de gênero. A utilização coletiva de ferramentas baseadas em inteligência artificial para produzir, compartilhar e legitimar conteúdos abusivos reforça mecanismos de controle e disciplinamento dos corpos femininos, especialmente por meio da exploração e da exposição não consentida da sexualidade.

Dessa forma, a misoginia colaborativa evidencia que a inteligência artificial não cria novas formas de dominação, mas potencializa estruturas patriarcais já existentes, reafirmando

o controle da sexualidade feminina como um dos elementos centrais da violência de gênero, conforme discutido anteriormente neste trabalho.

Portanto, a amplificação não deve ser atribuída integralmente à IA. Ela é direta quando a IA automatiza ataques, multiplica desinformações, recomenda conteúdos abusivos ou reproduz preconceitos em grande escala; indireta quando aumenta o volume de materiais disponíveis; e externa quando a circulação, o armazenamento e a permanência da violência dependem das plataformas e das comunidades de usuários. Essa distinção permite identificar com maior precisão a participação da IA sem confundir a produção tecnológica do conteúdo com sua circulação social.

4.4 USO DA IA NO ENFRENTAMENTO DA VIOLÊNCIA DE GÊNERO DIGITAL

A literatura analisada mostra que a IA pode auxiliar no enfrentamento e na proteção contra a violência de gênero digital. Neste trabalho, o enfrentamento compreende o uso da IA para identificar conteúdos e comportamentos abusivos, apoiar a moderação e reduzir sua disseminação. A proteção envolve orientar vítimas, detectar ameaças e facilitar o acesso a serviços de assistência. Foram consideradas apenas medidas com participação direta da IA.

Uma das aplicações mais desenvolvidas é a identificação automática de misoginia, sexismo, discurso de ódio e outros abusos textuais. Sistemas de Processamento de Linguagem Natural (PLN), aprendizado de máquina e aprendizado profundo analisam grandes volumes de publicações e sinalizam conteúdos possivelmente abusivos para moderação. Hashmi *et al.* (2025) desenvolveram um modelo para detectar mensagens misóginas em inglês e italiano utilizando três bases públicas com publicações do Twitter/X, Facebook e Reddit. A melhor configuração, composta pelo classificador FastText supervisionado e pelas redes neurais CNN-LSTM, alcançou 95% de precisão, *recall* e acurácia, além de F1-score de 0,95.

Os autores também utilizaram a técnica de IA Explicável LIME para indicar as palavras que mais influenciaram a classificação. Em alguns exemplos, termos como "*womensuck*", "*ruined*", "*bitch*", "*dirty*", "*hoe*" e "*cunt*" foram considerados relevantes para identificar mensagens misóginas. Essa explicação permite que os moderadores compreendam por que uma publicação foi sinalizada. Por isso, os autores defendem a integração entre IA e moderação humana.

A revisão de Neubauer *et al.* (2023) mostra que essas aplicações também podem identificar casos e tipos de violência, pedidos de ajuda, sentimentos, temas recorrentes e riscos de reincidência. Dos 22 estudos analisados, quinze utilizaram dados de redes sociais como Twitter/X, Reddit, Facebook e Baidu. Os modelos incluíram métodos supervisionados, sistemas baseados em regras, redes neurais e Transformers. Alguns sistemas baseados em regras alcançaram F1-score de 0,89 na identificação dos tipos de abuso e acurácia superior a 0,89 na detecção de violência doméstica. Além do desempenho, esses sistemas permitem compreender melhor as regras utilizadas. Contudo, como os estudos possuíam bases e objetivos diferentes,

seus resultados não permitem uma comparação direta.

No enfrentamento de *deepfakes* sexualmente explícitos, a visão computacional pode analisar imagens e vídeos em busca de inconsistências visuais, alterações faciais e outros sinais de manipulação. Maras e Logie (2024) explicam que, após a detecção, o material pode ser sinalizado, identificado visualmente, encaminhado para análise ou impedido de circular. Entretanto, o aprimoramento dos sistemas generativos torna os conteúdos falsos cada vez mais realistas, criando uma disputa contínua entre produção e detecção.

Maras e Logie (2024) também apresentam marcas d'água digitais, assinaturas, identificadores únicos e impressões digitais como formas de reconhecer a origem dos conteúdos e bloquear materiais manipulados. Esses registros poderiam formar uma biblioteca compartilhada de arquivos abusivos. Antes de aceitar uma publicação, a plataforma compararia automaticamente o arquivo com os registros existentes, bloqueando-o ou encaminhando-o para revisão em caso de correspondência. A eficácia da medida dependeria da integração entre os serviços e da capacidade de reconhecer versões modificadas do mesmo conteúdo.

A prevenção também pode ser incorporada ao desenvolvimento dos sistemas por meio do princípio de *ethics by design*, discutido por Maras e Logie (2024). Isso inclui bloquear pedidos de "nudificação", detectar tentativas de produzir material sexual sem consentimento, restringir a manipulação de imagens de pessoas reais e impedir conteúdos envolvendo crianças. Contudo, essas barreiras precisam interpretar a finalidade das solicitações, pois usuários podem reformular comandos ou empregar técnicas para contornar as restrições. A prevenção exige, portanto, atualização contínua e análise tanto das entradas quanto das respostas geradas, embora não consiga impedir totalmente o uso de modelos sem restrições ou desenvolvidos para fins abusivos.

Outra aplicação são os *chatbots* destinados à orientação e à proteção das vítimas. Imran e Tariq (2025) explicam que essas ferramentas podem oferecer atendimento anônimo, indicar canais de denúncia, orientar a preservação de provas, auxiliar na elaboração de planos de segurança e encaminhar a vítima para apoio jurídico ou especializado. Os autores mencionam a ferramenta Sara, criada para apoiar vítimas na América Central. Ganguly, Priya e Pandey (2026) também identificaram *chatbots* e ferramentas de aconselhamento virtual voltados à orientação, ao planejamento de segurança e ao apoio psicológico inicial. Essas aplicações podem ampliar o acesso à assistência onde os serviços presenciais são limitados. Entretanto, como o estudo é uma revisão qualitativa e não testa diretamente as ferramentas, ele apresenta possibilidades de uso sem comprovar que todas tenham a mesma eficácia.

A IA também pode avaliar riscos. Imran e Tariq (2025) mencionam modelos de aprendizado de máquina capazes de reconhecer padrões relacionados a vítimas, agressores e situações de violência. Ganguly, Priya e Pandey (2026) acrescentam que algoritmos podem analisar grandes volumes de publicações para identificar ameaças e ataques coordenados. Embora permitam perceber situações graves mais rapidamente, esses sistemas não devem decidir sozinhos se uma pessoa representa risco ou se uma vítima precisa de determinada

intervenção, sendo necessária a participação humana.

Essas ferramentas ainda possuem limitações. Os modelos identificam com maior facilidade insultos explícitos, mas apresentam dificuldades com ironias, sarcasmo, humor, sexismo benevolente e outras manifestações indiretas. Ganguly, Priya e Pandey (2026) observam ainda que sistemas treinados com dados da população majoritária podem não reconhecer formas específicas de violência contra pessoas LGBTQ+, minorias linguísticas e outros grupos marginalizados.

Os estudos convergem ao reconhecer o potencial da IA para identificar misoginia, discurso de ódio e conteúdos manipulados. Entretanto, as evidências não possuem o mesmo nível de comprovação. Enquanto Hashmi *et al.* (2025) testaram experimentalmente modelos e apresentaram métricas de desempenho, Imran e Tariq (2025) e Ganguly, Priya e Pandey (2026) discutem aplicações possíveis a partir de revisões da literatura. Portanto, a existência de possibilidades tecnológicas não significa que todas as ferramentas citadas tenham eficácia comprovada em situações reais. Também permanecem lacunas relacionadas ao reconhecimento de manifestações indiretas de violência, à diversidade linguística e à avaliação dos efeitos dessas tecnologias sobre as vítimas.

Em síntese, a IA pode contribuir para identificar conteúdos e comportamentos abusivos, impedir ou reduzir sua produção e orientar vítimas por meio de PLN, visão computacional, marcas e impressões digitais e *chatbots*. Entretanto, erros, vieses e limitações de contexto tornam indispensável a supervisão humana.

4.5 ASPECTOS JURÍDICOS, INSTITUCIONAIS E SOCIAIS DA VIOLÊNCIA DE GÊNERO DIGITAL

Além dos usos da IA na produção ou no enfrentamento da violência de gênero digital, os estudos analisados apresentam aspectos jurídicos, institucionais e sociais que ajudam a compreender o fenômeno.

No âmbito jurídico, Polyzoidou (2024) argumenta que a violência digital mantém relações com formas tradicionais de controle, humilhação e desigualdade contra as mulheres, mas apresenta características próprias, como anonimato, velocidade de disseminação, alcance internacional e reprodução contínua dos conteúdos. A autora relaciona o abuso de imagens íntimas e o discurso de ódio baseado em gênero à violação da dignidade; o *cyberstalking*, o *cyberharassment* e o *cyberbullying* à liberdade pessoal; e a exposição de conteúdos íntimos ou informações sensíveis à privacidade e à proteção dos dados pessoais. No Brasil, a Lei nº 15.123/2025 (Brasil, 2025) aumentou pela metade a pena do crime de violência psicológica contra a mulher quando praticado com IA ou outro recurso capaz de alterar sua imagem ou seu som. Entretanto, a aplicação da norma depende da presença dos elementos desse crime, não abrangendo automaticamente todo conteúdo íntimo artificial criado ou divulgado sem consentimento.

O consentimento ocupa posição central nessa discussão. A violência pode ocorrer tanto pela divulgação de uma imagem verdadeira quanto pela transformação de fotografias comuns ou criação de representações íntimas falsas. Por isso, a proteção deve considerar o uso da imagem sem autorização, e não apenas a autenticidade do conteúdo. Maras e Logie (2024) entendem os *deepfakes* nus ou sexualmente explícitos como abuso sexual baseado em imagens, pois sua criação pode violar a dignidade, a privacidade e a autonomia da pessoa retratada. Bellomo (2025) observa que muitas legislações se concentram na divulgação, embora o dano possa começar na fabricação da imagem, que pode ser utilizada para ameaçar, constranger ou chantagear a vítima. A Diretiva Europeia 2024/1385 (União Europeia, 2024a) contempla representações artificiais de pessoas em atividades sexuais e ameaças de divulgação, mas pode não abranger claramente todas as imagens de nudez sintética produzidas por aplicativos de "nudificação". Maras e Logie (2024) defendem ainda que as legislações alcancem conteúdos totalmente sintéticos envolvendo crianças, mesmo sem uma criança real no material original.

A dimensão internacional também dificulta a responsabilização, pois o conteúdo pode ser produzido, publicado e armazenado em países diferentes. Polyzoidou (2024) defende definições jurídicas mais compatíveis entre os países para facilitar a cooperação e evitar que uma conduta seja criminalizada em um território, mas permaneça sem tratamento específico em outro.

A responsabilidade não deve recair apenas sobre quem produz ou publica o conteúdo. Maras e Logie (2024) criticam a obrigação imposta à vítima de localizar cada cópia e solicitar individualmente sua remoção, pois o material pode continuar circulando em perfis, grupos, mecanismos de busca e serviços de armazenamento. Os autores defendem a retirada rápida e a responsabilização das plataformas e dos intermediários que não adotem providências após serem informados. No Brasil, o Decreto nº 12.976/2026 (Brasil, 2026) determina que as plataformas indisponibilizem, em até duas horas após a notificação, conteúdos íntimos exibidos sem autorização, incluindo os produzidos ou manipulados por IA. A retirada deve alcançar toda a aplicação, e o conteúdo deve ser marcado digitalmente para bloquear automaticamente seu reenvio. A norma também exige canais permanentes de denúncia e salvaguardas nos serviços de IA para identificar e bloquear solicitações de geração de conteúdo íntimo de terceiros.

Os avanços legislativos discutidos também permitem retomar a lacuna anteriormente identificada na Lei Maria da Penha, onde sua redação original não abrange as diversas manifestações da violência praticada em ambientes digitais. A promulgação da Lei nº 15.123/2025 e do Decreto nº 12.976/2026 representa um avanço ao reconhecer o papel das tecnologias digitais e da inteligência artificial na prática dessas violências, fortalecendo mecanismos de proteção e responsabilização. Entretanto, tais instrumentos não eliminam integralmente as limitações do ordenamento jurídico. Como observado neste trabalho, a ocorrência da Lei nº 15.123/2025 depende da configuração dos elementos específicos do crime previsto em lei, não abrangendo automaticamente toda forma de produção ou disseminação de conteúdo íntimo artificial. Assim, embora o novo marco normativo reduza parte da

lacuna anteriormente existente, permanecem desafios relacionados à abrangência das condutas tipificadas, à rápida evolução das tecnologias baseadas em inteligência artificial e à necessidade de constante atualização das políticas públicas e dos instrumentos jurídicos destinados ao enfrentamento da violência de gênero digital.

Bellomo (2025) estende essa responsabilidade aos desenvolvedores, fornecedores, empresas que comercializam os sistemas e plataformas que distribuem seus resultados. O Regulamento Europeu de IA (*AI Act*) (União Europeia, 2024b) estabelece obrigações de transparência, documentação, segurança e gerenciamento de riscos, enquanto o Regulamento dos Serviços Digitais (*Digital Services Act* — DSA) (União Europeia, 2022) exige que as plataformas avaliem e reduzam riscos sistêmicos. Contudo, identificar um conteúdo como artificial não resolve a ausência de consentimento nem elimina sua circulação e seus danos.

A atuação dos desenvolvedores foi investigada por Pawelec (2025) em entrevistas com dez especialistas em mídia sintética. Embora esses profissionais influenciem as funcionalidades, o acesso e os mecanismos de segurança das ferramentas, suas preocupações se concentraram em desinformação, fraudes e manipulação eleitoral. Privacidade, consentimento e pornografia *deepfake* receberam menor atenção, apesar de afetarem desproporcionalmente as mulheres.

Essa limitação revela uma fragmentação entre normas sobre IA, proteção de dados, serviços digitais e violência contra as mulheres. Bellomo (2025) recorre à Convenção sobre a Eliminação de Todas as Formas de Discriminação contra a Mulher (CEDAW) (Organização das Nações Unidas, 1979) e à Convenção de Istambul (Conselho da Europa, 2011) para defender uma resposta baseada na prevenção, na proteção das vítimas, na responsabilização dos agressores e em políticas integradas. Embora esses instrumentos não tenham sido elaborados especificamente para os *deepfakes*, seus princípios permitem reconhecer o abuso sexual baseado em imagens como violência de gênero, e não apenas como falsificação ou uso indevido da tecnologia.

Os impactos sociais e psicológicos reforçam essa necessidade. Na revisão de Diel *et al.* (2026), vítimas relataram medo, vergonha, culpa, depressão, isolamento, alterações no sono e dificuldades para trabalhar, entre outros efeitos. Contudo, dos 28 estudos incluídos, somente quatro examinaram diretamente os danos à saúde mental. Os resultados indicam a possível gravidade dos impactos, mas não permitem generalizá-los para todas as vítimas. Bellomo (2025) acrescenta que o medo de novas manipulações pode levar mulheres a excluir perfis, deixar de publicar fotografias ou reduzir sua participação nos ambientes digitais.

A prevenção também envolve educação midiática e mudanças no desenvolvimento das tecnologias. Diel *et al.* (2026) identificaram estudos nos quais programas educativos reduziram a intenção de compartilhar desinformação produzida por IA. Bellomo (2025) defende que a alfabetização digital explique como os *deepfakes* são produzidos, os riscos de compartilhá-los e as formas de denunciar conteúdos e preservar provas. Essa educação, porém, não deve transferir às mulheres a responsabilidade pela prevenção, mas complementar as obrigações do Estado, das empresas e das plataformas.

Em síntese, a violência de gênero digital mediada pela IA não é apenas um problema técnico, pois envolve direitos, consentimento, impactos psicológicos e responsabilidades distribuídas entre usuários, desenvolvedores, empresas, plataformas e governos. Nenhuma medida isolada é suficiente: criminalizar não garante a remoção, identificar um *deepfake* não corrige a ausência de consentimento e excluir uma publicação não encerra seus efeitos. Por isso, os estudos defendem uma resposta integrada que reúna legislação, responsabilização institucional, prevenção, educação e participação dos grupos afetados.

5 CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo analisar como a inteligência artificial tem sido utilizada tanto na produção e amplificação quanto no enfrentamento da violência de gênero contra a mulher no ambiente digital. A pesquisa permitiu compreender que a IA apresenta uma atuação ambígua: ao mesmo tempo que facilita novas formas de violência, também oferece recursos para identificá-las, preveni-las e combatê-las.

Os resultados demonstraram que a inteligência artificial não criou a violência de gênero digital, mas ampliou sua capacidade de produção e circulação. Ferramentas generativas permitem criar imagens, vídeos, áudios e textos falsos ou manipulados, como *deepfakes* sexuais e conteúdos de "nudificação", enquanto as plataformas digitais favorecem sua reprodução, permanência e disseminação. Dessa forma, práticas de controle, humilhação e objetificação já existentes passam a ocorrer com maior velocidade, alcance e facilidade técnica.

Por outro lado, técnicas de processamento de linguagem natural, aprendizado de máquina e visão computacional têm sido empregadas para detectar misoginia, discurso de ódio, assédio e conteúdos manipulados. Entretanto, essas ferramentas ainda apresentam limitações relacionadas aos vieses dos dados, à interpretação do contexto e aos erros de classificação. Por isso, não devem substituir a análise humana nem ser tratadas como soluções suficientes por si mesmas.

No campo jurídico-institucional, os resultados também evidenciaram que, embora o ordenamento brasileiro tenha avançado com a promulgação da Lei nº 15.123/2025 e do Decreto nº 12.976/2026, permanecem desafios relevantes para a proteção integral das vítimas de violência de gênero digital. As normas recentes representam um importante reconhecimento do papel da inteligência artificial na prática dessas violências e ampliam os instrumentos de enfrentamento, porém não eliminam a fragmentação normativa nem abrangem todas as situações envolvendo conteúdos íntimos gerados artificialmente. Conforme discutido, a caracterização de um conteúdo como produzido por inteligência artificial, por si só, não resolve questões centrais relacionadas à ausência de consentimento, à efetiva responsabilização dos agressores e à proteção das vítimas, evidenciando a necessidade de contínuo aperfeiçoamento legislativo, institucional e tecnológico.

Portanto, o enfrentamento da violência de gênero digital exige uma resposta conjunta.

Além das ferramentas tecnológicas, são necessárias medidas de prevenção, legislação adequada, responsabilização de agressores e plataformas, canais acessíveis de denúncia, apoio às vítimas e atenção ao consentimento durante o desenvolvimento e o uso dos sistemas de IA.

Como limitações, destaca-se o caráter bibliográfico e documental da pesquisa, sem aplicação ou avaliação direta de ferramentas de inteligência artificial. O recorte das bases, dos idiomas e do período analisado também restringiu o conjunto de estudos, enquanto a concentração da literatura nos *deepfakes* sexuais deixou outras manifestações da violência de gênero mediada por IA menos exploradas. Além disso, a própria interpretação dos estudos não permitiu assegurar a identificação exaustiva de todas as possíveis convergências e divergências presentes na literatura. Assim, determinadas nuances, perspectivas ou posicionamentos podem não ter sido percebidos ou incorporados à análise, ainda que não seja possível afirmar que tais elementos estejam necessariamente presentes nos estudos examinados.

Para trabalhos futuros, recomenda-se investigar essas formas de violência no contexto brasileiro, avaliar ferramentas de detecção em língua portuguesa e acompanhar a aplicação das normas recentes relacionadas à remoção de conteúdos e à responsabilização. Também são necessárias pesquisas que considerem as experiências das vítimas e as diferenças entre os grupos afetados.

REFERÊNCIAS

ALWIS, Rangita de Silva de. A rapidly shifting landscape: why digitized violence is the newest category of gender-based violence. **La Revue des Juristes de Sciences Po**, Paris, n. 25, p. 62, 2024. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4648409. Acesso em: 31 jan. 2026.

AMARAL, Pedro Henrique Borges Ribeiro. IMPORTUNAÇÃO SEXUAL: UMA ANÁLISE DA LEI Nº 13.718/2018. **RECIMA21 - Revista Científica Multidisciplinar**, [S. l.], v. 5, n. 5, p. e555224, 2024. DOI: 10.47820/recima21.v5i5.5224. Disponível em: <https://recima21.com.br/recima21/article/view/5224>. Acesso em: 2 fev. 2026.

ARAUJO, Maria de Fátima. Gênero e violência contra a mulher: o perigoso jogo de poder e dominação. **Psicol. Am. Lat.**, México, n. 14, out. 2008. Disponível em http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1870-350X2008000300012&lng=pt&nrm=iso. Acesso em: 20 jan. 2026.

BALAVADZE, Mariam. The Impact of Artificial Intelligence on Gender-Based Violence. **Perspectives on Federalism**, v. 17, n. 1, p. E-1–E-27, 2025. Disponível em: https://www.on-federalism.eu/images/2025/Vol17_issue1/PoF_2025_Vol17-1_1_Balavadze.pdf. Acesso em: 17 jul. 2026.

BANDEIRA, Lourdes Maria. Violência de gênero: a construção de um campo teórico e de investigação. **Sociedade e Estado**, v. 29, n. 2, p. 449–469, maio 2014.

BEAUVOIR, Simone. **O Segundo Sexo: Fatos e Mitos**. Tradução de Sérgio Millet. 4. Ed. São Paulo: Difusão Europeia do Livro, 1970.

BELLOMO, Giorgia. **Framing Gender-Based Violence in the Age of Artificial Intelligence: the Cases of Algorithmic Discrimination and of Deepfake Pornography**. Dissertação (Mestrado em Relações Internacionais Comparadas) – Università Ca' Foscari Venezia, Veneza, 2025. Disponível em: <https://unitesi.unive.it/handle/20.500.14247/28426>. Acesso em: 18 jul. 2026.

BRASIL. **Decreto nº 12.976, de 20 de maio de 2026**. Estabelece diretrizes para a proteção de mulheres na internet e para o enfrentamento da violência contra mulheres em ambiente digital. Diário Oficial da União, seção 1, Brasília, DF, ed. 94, p. 4, 21 maio 2026. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2026/decreto/d12976.htm. Acesso em: 22 jul. 2026.

BRASIL. **Lei nº 11.340, de 7 agosto de 2006**. Cria mecanismos para coibir a violência doméstica e familiar contra a mulher. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2006/lei/111340.htm. Acesso em: 19 de dez. 2025.

BRASIL. **Lei nº 12.737, de 30 de novembro de 2012**. Dispõe sobre a tipificação criminal de delitos informáticos. Brasília: Presidência da República, Casa Civil, Subchefia para Assuntos Jurídicos. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/112737.htm. Acesso em: 02 fev. 2026.

BRASIL. **Lei nº 13.718, de 24 de setembro de 2018**. Altera o Decreto-Lei nº 2.848, de 7 de dezembro de 1940 (Código Penal), para tipificar os crimes de importunação sexual e de divulgação de cena de estupro, tornar pública incondicionada a natureza da ação penal dos crimes contra a liberdade sexual e dos crimes sexuais contra vulnerável, estabelecer causas de aumento de pena para esses crimes e definir como causas de aumento de pena o estupro coletivo e o estupro corretivo. Diário Oficial da União, Brasília, 25 set. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113718.htm. Acesso em: 02 fev. 2026.

BRASIL. **Lei nº 15.123, de 24 de abril de 2025**. Altera o art. 147-B do Decreto-Lei nº 2.848, de 7 de dezembro de 1940 (Código Penal), para estabelecer causa de aumento de pena no crime de violência psicológica contra a mulher quando praticado com o uso de inteligência artificial ou de qualquer outro recurso tecnológico que altere imagem ou som da vítima. Diário Oficial da União, Brasília, DF, 25 abr. 2025. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2025/lei/115123.htm. Acesso em: 21 jul. 2026.

BRASIL. Senado Federal. DataSenado. **Pesquisa Nacional de Violência contra as Mulheres 2025**. Brasília, DF: Senado Federal, 2025. Disponível em: https://www.senado.leg.br/institucional/datasenado/relatorio_online/pesquisa_violencia_domestica/2025/interativo.html. Acesso em: 5 ago. 2026.

CONSELHO DA EUROPA. **Council of Europe Convention on Preventing and Combating Violence Against Women and Domestic Violence**. CETS nº 210, Istambul, 11 maio 2011. Disponível em: <https://rm.coe.int/168008482e>. Acesso em: 21 jul. 2026.

DEFENSORIA Pública do Estado do Ceará. **Lei Carolina Dieckmann**: 10 anos da lei que protege a privacidade dos brasileiros no ambiente virtual. Defensoria Pública do Estado do Ceará, 2022. Disponível em: <https://www.defensoria.ce.def.br/noticia/lei-carolina-dieckmann-10-anos-da-lei-que- protege-a-privacidade-dos-brasileiros-no-ambiente-virtual/>. Acesso em: 02 fev. 2026.

DIEL, Alexander; *et al.* The Harm of Deepfakes: A Scoping Review of Deepfakes' Negative Effects on Human Mind and Behavior. **AI & Society**, v. 41, p. 5971–5987, 2026. DOI: 10.1007/s00146-025-02774-0. Disponível em: <https://link.springer.com/article/10.1007/s00146-025-02774-0>. Acesso em: 20 jul. 2026.

DÖRING, Nicola; LE, Thuy Dung; MILLER, Dan J. Experiences with AI-Generated Pornography: A Quantitative Content Analysis of Reddit Posts. **Archives of Sexual Behavior**, 18 set. 2025. DOI: 10.1007/s10508-025-03227-x. Disponível em: <https://doi.org/10.1007/s10508-025-03227-x>. Acesso em: 22 jul. 2026.

DÖRING, Nicola; *et al.* The Impact of Artificial Intelligence on Human Sexuality: A Five-Year Literature Review 2020–2024. **Current Sexual Health Reports**, v. 17, art. 4, 2025. DOI: 10.1007/s11930-024-00397-y. Disponível em: <https://link.springer.com/article/10.1007/s11930-024-00397-y>. Acesso em: 18 jul. 2026.

DUTTA, Aditi; BANDUCCI, Susan; CAMARGO, Chico Q. Divided by Discipline? A Systematic Literature Review on the Quantification of Online Sexism and Misogyny Using a Semi-Automated Approach. **Scientometrics**, v. 130, p. 4915–4971, 2025. DOI: 10.1007/s11192-025-05410-2. Disponível em: <https://link.springer.com/article/10.1007/s11192-025-05410-2>. Acesso em: 19 jul. 2026.

FERRAZ, Thomas Palmeira. *et al.* Inteligência Artificial explicável para atenuar a falta de transparência e a legitimidade na moderação da Internet. **Estudos Avançados**, v. 38, n. 111, p. 381–405, 2024.

GANGULY, Tanni; PRIYA, Kanu; PANDEY, Aaratrika. Mitigating Socio-Demographic Disparities and Amplifying Justice: The Role of Artificial Intelligence in Addressing Technology Facilitated Gender-Based Violence in the South Asia. **Cadernos de Pós-Graduação em Direito Político e Econômico**, v. 26, n. S1, p. 214–228, 2026. Disponível em: <https://ceapress.org/index.php/cpgdpe/article/view/207>. Acesso em: 21 jul. 2026.

GOOGLE. **Inteligência artificial (IA): um guia fácil de entender**. Google Cloud. Disponível em: <https://cloud.google.com/learn/what-is-artificial-intelligence>. Acesso em : 06 fev. 2026.

HASHMI, Ehtesham; YAYILGAN, Sule Yildirim; YAMIN, Muhammad Mudassar; ULLAH, Mohib. Enhancing Misogyny Detection in Bilingual Texts Using Explainable AI and Multilingual Fine-Tuned Transformers. **Complex & Intelligent Systems**, v. 11, art. 39, 2025. DOI: 10.1007/s40747-024-01655-1. Disponível em: <https://doi.org/10.1007/s40747-024-01655-1>. Acesso em: 21 jul. 2026.

HOEGG, Emma. Technology-Facilitated Gender-Based Violence (TFGBV) and Generative Artificial Intelligence (AI). **That's What [We] Said**, Vancouver, v. 7, n. 1, 2025. Disponível em: <https://ojs-o.library.ubc.ca/index.php/thatswhatwesaid/article/view/622>. Acesso em: 20 jul. 2026.

IMRAN, Nazish; TARIQ, Khaula. Artificial Intelligence and Gender-Based Violence: A Double-Edged Sword in the Fight for Justice and Equality. **Annals of King Edward Medical University**, Lahore, v. 31, n. 1, p. 1–4, jan./mar. 2025. DOI: 10.21649/akemu.v31i1.6079. Disponível em: <https://doi.org/10.21649/akemu.v31i1.6079>. Acesso em: 22 jul. 2026.

KOPECKÝ, Kamil; VORÁČ, Dominik. The Phenomenon of Deep Nudes: A New Threat to Children and Adults. **AI & Society**, v. 41, p. 545–556, 2026. DOI: 10.1007/s00146-025-02425-4. Disponível em: <https://link.springer.com/article/10.1007/s00146-025-02425-4>. Acesso em: 21 jul. 2026.

LORCA, Jacinto G.; GARCÍA-MINGO, Elisa. “Busca, busca, perrita”: comunidades digitais misóginas de difusão de imagens sexuais sin consentimiento. **ex æquo**: Revista da Associação Portuguesa de Estudos sobre as Mulheres, n. 48, p. 15–32, 2023. DOI: 10.22355/exaequo.2023.48.03. Disponível em: <https://doi.org/10.22355/exaequo.2023.48.03>. Acesso em: 20 jul. 2026.

MARAS, Marie-Helen; LOGIE, Kenji. Countering the Complex, Multifaceted Nature of Nude and Sexually Explicit Deepfakes: An Augean Task? **Crime Science**, v. 13, art. 31, 2024. DOI: 10.1186/s40163-024-00226-6. Disponível em: <https://link.springer.com/article/10.1186/s40163-024-00226-6>. Acesso em: 19 jul. 2026.

NEUBAUER, Lilly; STRAW, Isabel; MARICONTI, Enrico; TANCZER, Leonie Maria. A Systematic Literature Review of the Use of Computational Text Analysis Methods in Intimate Partner Violence Research. **Journal of Family Violence**, v. 38, p. 1205–1224, 2023. DOI: 10.1007/s10896-023-00517-7. Disponível em: <https://doi.org/10.1007/s10896-023-00517-7>. Acesso em: 21 jul. 2026.

OFCOM (Office of Communications). **Online content moderation**. Ofcom, 2019. Disponível em: <https://www.ofcom.org.uk/online-safety/safety-technology/online-content-moderation>. Acesso em: 05 fev. 2026.

OPAS. **Violência contra as mulheres - OPAS/OMS | Organização Pan-Americana da Saúde**. 2020. Disponível em: <https://www.paho.org/pt/topics/violence-against-women>. Acesso em: 20 jan. 2026.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. **Convention on the Elimination of All Forms of Discrimination against Women**. Nova York, 18 dez. 1979. Disponível em: <https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-elimination-all-forms-discrimination-against-women>. Acesso em: 22 jul. 2026.

ORGANIZAÇÃO dos Estados Americanos (OEA). **Lei Modelo Interamericana para Prevenir, Punir e Erradicar a Violência Digital contra as Mulheres**. Fortaleza: Mecanismo de Acompanhamento da Convenção de Belém do Pará (MESECVI), 2025. Disponível em: <https://www.gov.br/mulheres/pt-br/assuntos/internacional/LeiModeloInteramericanaViolnciaDigitalcontraMulheres.pdf>. Acesso em: 28 jan. 2026.

PAWELEC, Maria. Decent Deepfakes? Professional Deepfake Developers’ Ethical Considerations and Their Governance Potential. **AI & Ethics**, v. 5, p. 2641–2666, 2025. DOI: 10.1007/s43681-024-00542-2. Disponível em: <https://doi.org/10.1007/s43681-024-00542-2>. Acesso em: 22 jul. 2026.

POLYZOIDOU, Vagia. Digital Violence Against Women: Is There a Real Need for Special Criminalization? **International Journal for the Semiotics of Law**, v. 37, n. 6, p. 1777–1797, 2024. DOI: 10.1007/s11196-024-10179-3. Disponível em: <https://link.springer.com/article/10.1007/s11196-024-10179-3>. Acesso em: 19 jul. 2026.

RACHMARIA, Laksmi; SUSANTO, Andy. Potensi Kekerasan Gender Berbasis Online pada Penyalahgunaan Teknologi Kecerdasan Buatan bagi Perempuan di Media. **Jurnal Netnografi Komunikasi**, v. 2, n. 2, p. 91–104, 2024. DOI: 10.59408/jnk.v2i2.25. Disponível em: <https://netnografiikom.org/index.php/netnografi/article/view/25>. Acesso em: 19 jul. 2026.

RASHID, Adib Bin; KAUSIK, Md Ashfakul Karim. **AI revolutionizing industries worldwide**: a comprehensive overview of its diverse applications. *Hybrid Advances*, v. 7, p. 100277, 2024. DOI: <https://doi.org/10.1016/j.hybadv.2024.100277>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2773207X24001386>. Acesso em: 08 fev. 2026.

RUSSELL, Stuart; NORVIG, Peter. **Inteligência Artificial**: Uma Abordagem Moderna. Tradução de Regina Célia Simille de Macedo. 3 ed. Rio de Janeiro: Elsevier Editora, 2013.

SAFFIOTI, Heleieth Iara Bongiovani. **Gênero patriarcado violência**. 2. ed. São Paulo: Expressão Popular, 2015.

STRYKER, Cole; KAVLAKOGLU, Eda. **What is artificial intelligence (AI)?** IBM Think. 2024. Disponível em: <https://www.ibm.com/think/topics/artificial-intelligence>. Acesso em: 05 fev. 2026.

UNIÃO EUROPEIA. **Diretiva (UE) 2024/1385 do Parlamento Europeu e do Conselho, de 14 de maio de 2024**. Relativa ao combate à violência contra as mulheres e à violência doméstica. *Jornal Oficial da União Europeia*, 24 maio 2024. Disponível em: <https://eur-lex.europa.eu/eli/dir/2024/1385/oj/por>. Acesso em: 21 jul. 2026.

UNIÃO EUROPEIA. **Regulamento (UE) 2022/2065 do Parlamento Europeu e do Conselho, de 19 de outubro de 2022**. Relativo a um mercado único para os serviços digitais e que altera a Diretiva 2000/31/CE (Regulamento dos Serviços Digitais). *Jornal Oficial da União Europeia*, L 277, p. 1–102, 27 out. 2022. Disponível em: <https://eur-lex.europa.eu/eli/reg/2022/2065/oj/por>. Acesso em: 21 jul. 2026.

UNIÃO EUROPEIA. **Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024**. Cria regras harmonizadas em matéria de inteligência artificial e altera determinados atos legislativos da União Europeia (Regulamento da Inteligência Artificial). *Jornal Oficial da União Europeia*, 12 jul. 2024. Disponível em: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/por>. Acesso em: 20 jul. 2026.