VI CONGRESSO INTERNACIONAL DE DIREITO E INTELIGÊNCIA ARTIFICIAL (VI CIDIA)

REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL II

R344
Regulação da inteligência artificial II [Recurso eletrônico on-line] organização VI Congresso Internacional de Direito e Inteligência Artificial (VI CIDIA): Skema Business School – Belo Horizonte;

Coordenadores: Danúbia Patrícia de Paiva, David França Carvalho e Renata Kretzmann – Belo Horizonte: Skema Business School, 2025.

Inclui bibliografia
ISBN: 978-65-5274-354-1
Modo de acesso: www.conpedi.org.br em publicações
Tema: Perspectivas globais para a regulação da inteligência artificial.
1. Compliance. 2. Ética. 3. Legislação. I. VI Congresso Internacional de Direito e Inteligência Artificial (1:2025 : Belo Horizonte, MG).

CDU: 34

VI CONGRESSO INTERNACIONAL DE DIREITO E INTELIGÊNCIA ARTIFICIAL (VI CIDIA)

REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL II

Apresentação

A SKEMA Business School é uma organização francesa sem fins lucrativos, com presença em sete países diferentes ao redor do mundo (França, EUA, China, Brasil, Emirados Árabes Unidos, África do Sul e Canadá) e detentora de três prestigiadas acreditações internacionais (AMBA, EQUIS e AACSB), refletindo seu compromisso com a pesquisa de alta qualidade na economia do conhecimento. A SKEMA reconhece que, em um mundo cada vez mais digital, é essencial adotar uma abordagem transdisciplinar.

Cumprindo esse propósito, o VI Congresso Internacional de Direito e Inteligência Artificial (VI CIDIA), realizado nos dias 18 e 19 de setembro de 2025, em formato híbrido, manteve-se como o principal evento acadêmico sediado no Brasil com o propósito de fomentar ricas discussões sobre as diversas interseções entre o direito e a inteligência artificial. O evento, que teve como tema central a "Regulação da Inteligência Artificial", contou com a presença de renomados especialistas nacionais e internacionais, que abordaram temas de relevância crescente no cenário jurídico contemporâneo.

Profissionais e estudantes dos cursos de Direito, Administração, Economia, Ciência de Dados, Ciência da Computação, entre outros, tiveram a oportunidade de se conectar e compartilhar conhecimentos, promovendo um ambiente de rica troca intelectual. O VI CIDIA contou com a participação de acadêmicos e profissionais provenientes de diversas regiões do Brasil e do exterior. Entre os estados brasileiros representados, estavam: Alagoas (AL), Bahia (BA), Ceará (CE), Goiás (GO), Maranhão (MA), Mato Grosso do Sul (MS), Minas Gerais (MG), Pará (PA), Paraíba (PB), Paraná (PR), Pernambuco (PE), Piauí (PI), Rio de Janeiro

Foram discutidos assuntos variados, desde a própria regulação da inteligência artificial, eixo central do evento, até as novas perspectivas de negócios e inovação, destacando como os algoritmos estão remodelando setores tradicionais e impulsionando a criação de empresas inovadoras. Com uma programação abrangente, o congresso proporcionou um espaço vital para discutir os desafios e oportunidades que emergem com o desenvolvimento algorítmico, reforçando a importância de uma abordagem jurídica e ética robusta nesse contexto em constante evolução.

A programação teve início às 13h, com o check-in dos participantes e o aquecimento do público presente. Às 13h30, a abertura oficial foi conduzida pela Prof.ª Dr.ª Geneviève Poulingue, que, em sua fala de boas-vindas, destacou a relevância do congresso para a agenda global de inovação e o papel da SKEMA Brasil como ponte entre a academia e o setor produtivo.

Em seguida, às 14h, ocorreu um dos momentos mais aguardados: a Keynote Lecture do Prof. Dr. Ryan Calo, renomado especialista internacional em direito e tecnologia e professor da University of Washington. Em uma conferência instigante, o professor explorou os desafios metodológicos da regulação da inteligência artificial, trazendo exemplos de sua atuação junto ao Senado dos Estados Unidos e ao Bundestag alemão.

A palestra foi seguida por uma sessão de comentários e análise crítica conduzida pelo Prof. Dr. José Luiz de Moura Faleiros Júnior, que contextualizou as reflexões de Calo para a realidade brasileira e fomentou o debate com o público. O primeiro dia foi encerrado às 14h50 com as considerações finais, deixando os participantes inspirados para as discussões do dia seguinte.

As atividades do segundo dia tiveram início cedo, com o check-in às 7h30. Às 8h20, a Prof.ª Dr.ª Margherita Pagani abriu a programação matinal com a conferência Unlocking Business

Após um breve e merecido coffee break às 9h40, os participantes retornaram para uma manhã de intensas reflexões. Às 10h30, o pesquisador Prof. Dr. Steve Ataky apresentou a conferência Regulatory Perspectives on AI, compartilhando avanços e desafios no campo da regulação técnica e ética da inteligência artificial a partir de uma perspectiva global.

Encerrando o ciclo de palestras, às 11h10, o Prof. Dr. Filipe Medon trouxe ao público uma análise profunda sobre o cenário brasileiro, com a palestra AI Regulation in Brazil. Sua exposição percorreu desde a criação do Marco Legal da Inteligência Artificial até os desafios atuais para sua implementação, envolvendo aspectos legislativos, econômicos e sociais.

Nas tardes dos dois dias, foram realizados grupos de trabalho que contaram com a apresentação de cerca de 60 trabalhos acadêmicos relacionados à temática do evento. Com isso, o evento foi encerrado, após intensas discussões e troca de ideias que estabeleceram um panorama abrangente das tendências e desafios da inteligência artificial em nível global.

Os GTs tiveram os seguintes eixos de discussão, sob coordenação de renomados especialistas nos respectivos campos de pesquisa:

a) Startups e Empreendedorismo de Base Tecnológica – Coordenado por Allan Fuezi de Moura Barbosa, Laurence Duarte Araújo Pereira, Cildo Giolo Júnior, Maria Cláudia Viana Hissa Dias do Vale Gangana e Yago Oliveira

b) Jurimetria Cibernética Jurídica e Ciência de Dados – Coordenado por Arthur Salles de Paula Moreira, Gabriel Ribeiro de Lima, Isabela Campos Vidigal Martins, João Victor Doreto e Tales Calaza

c) Decisões Automatizadas e Gestão Empresarial / Algoritmos, Modelos de Linguagem e Propriedade Intelectual – Coordenado por Alisson Jose Maia Melo, Guilherme Mucelin e

f) Regulação da Inteligência Artificial – III – Coordenado por Ana Júlia Silva Alves Guimarães, Erick Hitoshi Guimarães Makiya, Jessica Fernandes Rocha, João Alexandre Silva Alves Guimarães e Luiz Felipe Vieira de Siqueira

g) Inteligência Artificial, Mercados Globais e Contratos – Coordenado por Gustavo da Silva Melo, Rodrigo Gugliara e Vitor Ottoboni Pavan

h) Privacidade, Proteção de Dados Pessoais e Negócios Inovadores – I – Coordenado por Dineia Anziliero Dal Pizzol, Evaldo Osorio Hackmann, Gabriel Fraga Hamester, Guilherme Mucelin e Guilherme Spillari Costa

i) Privacidade, Proteção de Dados Pessoais e Negócios Inovadores – II – Coordenado por Alexandre Schmitt da Silva Mello, Lorenzzo Antonini Itabaiana, Marcelo Fonseca Santos, Mariana de Moraes Palmeira e Pietra Daneluzzi Quinelato

j) Empresa, Tecnologia e Sustentabilidade – Coordenado por Marcia Andrea Bühring, Ana Cláudia Redecker, Jessica Mello Tahim e Maraluce Maria Custódio.

Cada GT proporcionou um espaço de diálogo e troca de experiências entre pesquisadores e profissionais, contribuindo para o avanço das discussões sobre a aplicação da inteligência artificial no direito e em outros campos relacionados.

Um sucesso desse porte não seria possível sem o apoio institucional do Conselho Nacional de Pesquisa e Pós-graduação em Direito - CONPEDI, que desde a primeira edição do evento provê uma parceria sólida e indispensável ao seu sucesso. A colaboração contínua do CONPEDI tem sido fundamental para a organização e realização deste congresso, assegurando a qualidade e a relevância dos debates promovidos.

Reitora – SKEMA Business School - Campus Belo Horizonte

Prof. Ms. Dorival Guimarães Pereira Júnior

Coordenador do Curso de Direito – SKEMA Law School

Prof. Dr. José Luiz de Moura Faleiros Júnior

Coordenador de Pesquisa – SKEMA Law School

INTELIGÊNCIA ARTIFICIAL, DEEPFAKES E DANOS À PERSONALIDADE: NOVOS DESAFIOS À TUTELA JURÍDICA DA IMAGEM E DA HONRA

ARTIFICIAL INTELLIGENCE, DEEPFAKES AND PERSONALITY DAMAGE: NEW CHALLENGES FOR THE LEGAL PROTECTION OF IMAGE AND HONOUR

José Luiz de Moura Faleiros Júnior [1]
Paula Bittencourt Carvalho [2]
Julia Eleuterio Oliveira Coimbra [3]

Resumo

Deepfakes, fruto de sistemas de inteligência artificial generativa, baratearam de forma inédita a produção de conteúdos audiovisuais hiper-realistas capazes de usurpar identidades, macular reputações e violar a privacidade. Este artigo investiga como o Direito Civil brasileiro pode aperfeiçoar a tutela dos direitos da personalidade — honra, imagem e intimidade — diante desses danos sintéticos. Após mapear aspectos técnicos e vetores de risco, analisamos legislação, jurisprudência e experiências comparadas, apontando lacunas interpretativas. Propomos critérios dogmáticos, instrumentos processuais de urgência e deveres de plataforma que conciliem inovação e dignidade humana no país.

Palavras-chave: Deepfakes, Inteligência artificial, Direitos da personalidade, Direito de imagem, Responsabilidade civil

Abstract/Resumen/Résumé

Deepfake technology, powered by generative artificial intelligence, has dramatically lowered the cost of fabricating hyper-realistic audiovisual content that can misappropriate identity, distort reputation and invade privacy. This article examines how Brazilian civil law can update the traditional protection of personality rights—honour, image and intimacy—to face such synthetic harms. After mapping technical features and threat vectors of deepfakes, we analyse relevant statutes, case law and comparative regulation, highlighting interpretative gaps. We propose doctrinal criteria and procedural tools for liability, preventive injunctions and platform duties, aligning innovation with human-dignity safeguards today.

1. INTRODUÇÃO

A difusão dos sistemas de IA generativa inaugurou uma fase em que qualquer usuário conectado pode produzir vídeos, áudios ou fotografias hiper-realistas capazes de replicar feições, gestos e voz de terceiros com precisão quase imperceptível. Os *deepfakes* deixam de ser curiosidade tecnológica para se tornarem vetor de riscos reputacionais, eleitorais e econômicos.

Tais artefatos sintéticos desafiam categorias jurídicas consolidadas: manipulam a percepção pública, erodem a confiança em registros audiovisuais e criam externalidades negativas que ultrapassam a esfera individual, afetando até a estabilidade de instituições democráticas.

No plano da dogmática civil, honra, imagem e privacidade compõem o núcleo duro dos direitos da personalidade, protegidos pela Constituição de 1988, pelos arts. 11-21 do Código Civil e por estatutos específicos (Marco Civil da Internet, LGPD). Entretanto, esses diplomas foram concebidos para mídias tradicionais e não previram a escalabilidade algorítmica da falsificação.

A jurisprudência nacional já tratava de fotomontagens e boatos digitais, mas os *deepfakes* introduzem novos dilemas: reversão do ônus da prova, quantificação de danos morais difusos, rapidez necessária às tutelas inibitórias e responsabilidade solidária entre criadores, provedores de IA e plataformas de hospedagem.

No exterior, iniciativas como o EU AI Act (2024) e a Seção 230 Reform Act (EUA) apontam caminhos para regulação de conteúdos sintéticos, exigindo rotulagem, rastreabilidade e avaliação de risco. O Brasil discute propostas afins no PL 2630/2020 e em projetos sobre desinformação, mas sem enfoque específico na violação de personalidade.

A literatura acadêmica, embora crescente, ainda carece de estudo sistemático que integre análise tecnológica, teoria dos direitos da personalidade e mecanismos processuais aptos a conter os danos *ex ante* — especialmente num país de forte cultura de redes sociais e judicialização.

Diante desse cenário, este artigo pergunta: de que modo o Direito Civil, especialmente a tutela dos direitos da personalidade, pode enfrentar os desafios trazidos pelos *deepfakes* e demais conteúdos sintéticos produzidos por IA? Para responder, desenvolve-se investigação que combina abordagem dogmática e análise comparada, estruturada em quatro seções além desta introdução.

Tem-se o objetivo de avaliar a suficiência do arcabouço jurídico brasileiro na proteção da honra, imagem e privacidade contra *deepfakes*, identificando lacunas e propondo soluções normativas e procedimentais que assegurem reparação célere e efetiva, bem como incentivos à prevenção.

2. CONTEÚDOS SINTÉTICOS E DANOS À PERSONALIDADE

O dano à personalidade caracteriza-se como lesão a atributos essenciais da pessoa, compreendendo honra objetiva, honra subjetiva, imagem e vida privada. *Deepfakes* potencializam tais lesões ao simular comportamentos desabonadores capazes de comprometer reputação em segundos. (Silveira; Faleiros Júnior, 2024) A verossimilhança técnica amplia a credibilidade do conteúdo e dificulta a prova de manipulação. A vítima enfrenta ônus informacional desproporcional, pois precisa de perícia avançada para demonstrar a falsidade. Surge, portanto, a necessidade de inversão dinâmica do ônus probatório. Esse deslocamento alinha-se à lógica protetiva dos direitos da personalidade. (Medon, 2021)

O reconhecimento do *deepfake* como ilícito pressupõe análise de ilicitude formal e material. Há ilicitude formal quando o autor viola deveres de veracidade e atribui falsamente comportamento à vítima. A ilicitude material revela-se pelo impacto negativo na esfera moral ou econômica do atingido. A aferição do nexo causal entre conteúdo e dano exige raciocínio contrafactual adaptado à viralidade das redes. (Silveira; Auto; Faleiros Júnior, 2024) A doutrina tem defendido adoção da teoria da causalidade probabilística, substituindo a certeza empírica pela alta plausibilidade. Tal abordagem reduz assimetria informacional.

A quantificação do dano moral em *deepfakes* exige critérios agravados pela perenidade digital. A perpetuidade do arquivo e a replicabilidade ilimitada conduzem ao que a literatura denomina dano continuado. (Medon, 2021) O magistrado deve ponderar extensão temporal, amplitude de difusão e intensidade de humilhação. Propõe-se que tais vetores componham um coeficiente multiplicador a ser aplicado na indenização. Essa técnica permitiria respostas mais aderentes ao impacto real. Também favoreceria previsibilidade, reforçando segurança jurídica.

No plano coletivo, *deepfakes* podem atingir grupos difusos, abalando confiança em instituições. Quando o conteúdo envolve agentes políticos, o dano extrapola a esfera individual e alcança a integridade do processo democrático. Doutrina e jurisprudência admitem dano moral coletivo sempre que há transgressão a valores sociais relevantes. O Ministério Público e entidades de defesa de consumidores poderiam ajuizar ações civis públicas para coibir

circulação do *deepfake* e obter reparação em favor da sociedade. A indenização poderia destinar-se a fundos de direitos difusos. Desse modo, o sistema amplia alcance protetivo.

O direito comparado oferece pistas valiosas. O AI Act da União Europeia (União Europeia, 2024) estabelece dever de rotulagem inequívoca para conteúdos sintéticos e impõe obrigações de transparência a provedores de sistemas generativos. Esse marco normativo orienta Estados-membros a exigir marcas d'água digitais e metadados padronizados, facilitando perícia. A adoção desse modelo no Brasil mitigaria a dificuldade probatória. A jurisprudência poderia presumir culpa do agente que publica *deepfake* não rotulado. Tal presunção estimularia conformidade preventiva.

A Lei norte-americana Take It Down Act, promulgada em maio de 2025, foca em *deepfakes* íntimos não consentidos e estabelece procedimentos expeditos de remoção. (United States of America, 2025) Impõe sanções civis severas às plataformas que mantêm conteúdos após notificação válida. Prevê ainda subsidiariedade de ação contra criadores e facilitadores tecnológicos. O modelo inspira integração entre tutela civil e direito administrativo sancionador. Sua lógica demonstra a eficácia de esquemas híbridos de *enforcement*. O transplante adaptado ao Brasil pode reforçar tutela da dignidade sexual.

O PL 2630/2020, originalmente conhecido como "Projeto de Lei das Fake News", tramita no Congresso Nacional e inclui obrigações de rastreabilidade e contas verificadas. (Brasil, 2020) Embora o texto ainda não trate especificamente de *deepfakes*, oferece infraestrutura normativa para futura inclusão de artigos sobre conteúdos sintéticos. Sugere-se inserir dispositivos que obriguem identificação de mídia manipulada, bem como janelas de contexto informacional. Tal atualização harmonizaria o projeto às tendências estrangeiras. A lacuna atual representa oportunidade para aperfeiçoamento legislativo imediato. A inércia pode agravar danos em ano eleitoral.

O elemento subjetivo do criador do *deepfake*, embora relevante para eventual dano punitivo, não condiciona o dever reparatório. A responsabilidade civil brasileira admite forma objetiva quando o risco da atividade é exacerbado. Gerar vídeo falso em escala ampla configura atividade de risco, atraindo a cláusula geral do art. 927, parágrafo único, do Código Civil. Assim, a vítima não precisaria demonstrar culpa, bastando comprovar evento, autoria e nexo causal. Esse deslocamento prestigia a função preventiva da responsabilidade. Converge com teorias de custo-benefício adotadas em regimes estrangeiros.

A defesa do *deepfake* muitas vezes invoca liberdade de expressão e direito de paródia. Entretanto, tais liberdades encontram limites no princípio da dignidade da pessoa humana e na vedação ao anonimato. O balanceamento de direitos fundamentais deve observar

proporcionalidade e necessidade. A sátira política permanece protegida desde que claramente identificável como ficção. Quando a técnica obscurece a natureza ficcional, a tutela da personalidade prevalece. O ônus de explicitar a paródia recairá sobre o criador.

Em matéria eleitoral, diversos estados dos EUA proíbem *deepfakes* enganosos nos 90 dias que antecedem o pleito. A experiência comparada demonstra que janelas temporais específicas podem mitigar manipulação de voto. (Minnesota, 2023) O Brasil, com histórico de judicialização eleitoral, poderia adotar regra análoga via resolução do TSE. A medida evitaria inundação de vídeos falsos às vésperas das urnas. Complementarmente, plataformas seriam obrigadas a criar canais prioritários de denúncia. Tais instrumentos reforçam a lisura democrática. (Mulholland; Oliveira, 2021)

O dano emergente decorrente da remoção tardia de *deepfake* inclui gastos com monitoramento, assessoria de imprensa e suporte psicológico. Tais custos devem ser ressarcidos sob a rubrica de danos materiais. (Ziobroń, 2024) A vítima frequentemente contrata empresas de monitoramento de imagem para varrer replicações. Esse desembolso guarda nexo direto com o ilícito. Reconhecer essa parcela indenizatória alinha-se ao princípio da *restitutio in integrum*. A reparação deve abranger todos os reflexos patrimoniais.

No campo penal, o art. 154-A do Código Penal, que tipifica invasão de dispositivo informático, poderia subsidiariamente enquadrar quem se apossa indevidamente de imagem alheia para criar *deepfake*. Todavia, a lacuna de tipicidade específica sugere necessidade de novo tipo penal. (Pinto; Oliveira, 2023) Doutrina penal econômica alerta para risco de expansão excessiva do direito criminal. Propõe-se, então, solução combinada: tipos penais mínimos e sanções civis efetivas. Essa dupla via assegura dissuasão sem hipertrofia punitivista. O debate legislativo encontra-se em fase embrionária.

A tutela inibitória apresenta-se como remédio processual central. A celeridade é crucial, pois cada minuto de permanência online amplia dano. Recomenda-se adoção de medidas liminares inaudita altera parte com base em verossimilhança da prova pericial inicial. A multa diária deve ser calibrada segundo capacidade econômica do réu e alcance da difusão. Medidas de *takedown* internacional podem demandar cooperação via Convenção de Budapeste. A jurisdição brasileira já admite ordens globais em casos de desinformação.

A teoria da responsabilidade civil gradada autoriza cumulação de indenização compensatória e punitiva em hipóteses de dolo específico ou finalidade comercial. O Superior Tribunal de Justiça ainda resiste à adoção de *punitive damages*, mas tem admitido majoração de danos morais em situações de reiteração. *Deepfakes* com motivação lucrativa, como pornografia de vingança, ajustam-se a tal lógica. A penalização pecuniária robusta desestimula

condutas futuras. Contribui para efeito pedagógico e satisfativo. A doutrina sinaliza convergência. (Geng, 2023)

Em síntese, o dano à personalidade causado por *deepfake* reclama interpretação evolutiva do direito civil. A dogmática deve acolher critérios tecnológicos, redistribuir ônus probatório e prever mecanismos de cálculo indenizatório sensíveis à perenidade digital. A teoria do risco da atividade sustenta responsabilidade objetiva para quem utiliza modelos generativos. Iniciativas estrangeiras de rotulagem e janelas eleitorais oferecem insumos regulatórios. O Brasil dispõe de base normativa, mas necessita ajustes cirúrgicos. O próximo capítulo examina responsabilidade dos intermediários digitais.

3. REGIMES DE RESPONSABILIDADE E DEVERES DE PRECAUÇÃO

A dinâmica dos *deepfakes* envolve cadeia complexa de atores que inclui desenvolvedor do modelo, usuário gerador, plataforma de hospedagem e mecanismo de busca. A atribuição de responsabilidade solidária encontra amparo no art. 942 do Código Civil quando há concurso de condutas. A solidariedade funda-se em unicidade do dano e pluralidade de causadores. Provedores de IA, ao disponibilizar modelos *open source*, criam risco inerente. Hospedagens que monetizam conteúdo falso beneficiam-se economicamente da projeção. Todos compartilham dever de prevenção.

O regime de responsabilidade das plataformas no Brasil ancora-se nos arts. 18 e 19 do Marco Civil da Internet. Em regra, vigora modelo de responsabilidade subjetiva condicionada a notificação judicial prévia. Contudo, a rapidez dos *deepfakes* torna esse filtro insuficiente. Propõe-se migração para sistema híbrido que combine notificação extrajudicial qualificada e responsabilidade objetiva para conteúdos manifestamente ilícitos. Essa solução preserva liberdade de expressão, mas impõe diligência proativa mínima. O precedente do tema 987 do STF indica espaço para diferenciação.

O AI Act europeu classifica sistemas que geram conteúdo sintético como “alto risco” e exige políticas de gestão de risco contínuo. (União Europeia, 2024) As plataformas devem implementar marca d’água robusta, metadados e sistemas de detecção automática. Descumprimento sujeita-se a multas administrativas que podem alcançar 6% do volume global de negócios. A internalização de custos incentiva compliance preventivo. Incorporar parâmetros semelhantes em regulamentação da ANPD reforçaria coerência com padrões internacionais. Evita-se *dumping* regulatório.

Provedores de hospedagem detêm controle técnico sobre remoção e bloqueio geográfico. O dever de retirada imediata após notificação fundamentada decorre do art. 927, parágrafo único, combinado com a boa-fé objetiva. A jurisprudência do STJ já reconheceu responsabilidade objetiva de aplicativos de mensagem em casos de abuso sexual infantil. O mesmo raciocínio aplica-se aos *deepfakes* por analogia. Argumenta-se que ausência de filtro mínimo viola dever anexo de cautela. A prevenção supera o mero reparo *ex post*.

Os modelos fundacionais podem incorporar salvaguardas como *refusal classes* para impedir geração de imagens íntimas não consentidas. Ferramentas de *hash* contra pornografia infantil funcionam como precedente técnico. Legislador pode exigir certificação de segurança algorítmica antes da liberação pública. A ANATEL poderia condicionar funcionamento de aplicativos nacionais a padrões de detecção automática. Esse licenciamento preventivo dialoga com ideia de *due diligence* algorítmica. Fomenta mercado de soluções de rastreabilidade.

A reformulação da Seção 230 nos Estados Unidos acena com redução de imunidades para plataformas que se beneficiem economicamente de *deepfakes* nocivos. (United States of America, 2023) A tendência global é atenuar blindagem jurídica de intermediários. No Brasil, a adoção de cláusula geral de cuidado pode ser inserida no PL 2630/2020 para responsabilizar provedores omissos. (Brasil, 2020) Essa medida encontra respaldo no princípio da cooperação processual e na função social da liberdade de expressão. Plataformas passariam a demonstrar cumprimento de protocolos de mitigação. O ônus de exculpação deslocar-se-ia em benefício da vítima. (De Ruiter, 2021)

Sistemas de *bounty*, nos quais pesquisadores recebem recompensa por detectar *deepfakes*, mostram-se eficazes para auditoria colaborativa. Regulamentos poderiam exigir destinação de percentual da receita publicitária a programas de *bug bounty* contra falsificação audiovisual. Essa abordagem amplia cobertura de detecção sem custos desproporcionais ao Estado. Paralelamente, universidades poderiam ser financiadas para desenvolver algoritmos *open source* de verificação. A articulação entre setor público, privado e academia conforma ecossistema de governança. Resulta em compliance distribuído. (Spivak, 2019)

A transparência algorítmica deve contemplar logs de geração e rastreabilidade de prompts. Esses registros permitem identificar usuário que criou conteúdo ilícito. A Lei Geral de Proteção de Dados Pessoais autoriza tratamento de dados pessoais quando necessário para exercício regular de direito. Assim, a guarda de logs atende legitimo interesse de responsabilização civil. Recomenda-se prazo mínimo de 180 dias, prorrogável por ordem judicial. Tal medida harmoniza privacidade e *accountability*.

A responsabilidade objetiva pode ser graduada por barreiras tecnológicas disponíveis. Plataformas que investem em detecção de *deepfake* e resposta rápida poderiam pleitear redução de danos punitivos. Prevê-se adoção de regime de "porto seguro condicional" condicionado à prova de esforços razoáveis. Esse mecanismo inspira o *Digital Services Act* europeu. (União Europeia, 20222) O incentivo econômico acelera adoção de boas práticas. Contribui para cultura de responsabilidade compartilhada. (Renaud, 2019)

Modelos de autorregulação merecem encorajamento. Códigos de conduta setoriais, validados por autoridade supervisora, permitem atualização dinâmica de padrões. A experiência da *EU Code of Practice on Disinformation* demonstra que compromissos voluntários podem reduzir propagação de falso conteúdo político. No entanto, sanções estatais permanecem essenciais para assegurar efetividade. O Brasil poderia instituir selo de conformidade para plataformas em conformidade com métricas de transparência. O benefício seria reputacional e regulatório.

A *due diligence* pré-lançamento de modelos generativos exige avaliação de impacto em direitos fundamentais. Inspirado no RGPD europeu, o relatório deveria ser público, possibilitando escrutínio de sociedade civil. Falhas detectadas poderiam implicar recall do modelo ou patch de segurança. A ausência de relatório acarretaria multa administrativa e responsabilidade solidária por danos. Tal instrumento antecipa riscos e reduz litigiosidade futura. Fomenta cultura de precaução.

Plataformas de pagamento online podem colaborar no estrangulamento financeiro de criadores de *deepfakes* ilícitos. Congelar repasses de monetização até verificação de autenticidade diminui incentivo econômico. A cooperação intersetorial amplia efetividade de medidas civis. Esse arranjo segue racionalidade das sanções bancárias contra streaming pirata. A responsabilização, portanto, estende-se à infraestrutura de pagamento. Harmoniza-se com teoria do facilitador indispensável.

A rotulagem obrigatória de conteúdo sintético, prevista no art. 50 do AI Act, estabelece padrão técnico mínimo de marcação em metadados. (União Europeia, 2024) Tal exigência facilita detecção por softwares de terceiros e serve como prova negativa em juízo. A ausência de rótulo gera presunção de ilicitude, invertendo ônus da prova. Transferir esse dispositivo para legislação brasileira reduziria litigiosidade probatória. Também fomentaria mercado de *watermarking* confiável. O alinhamento internacional beneficia fluxos transfronteiriços de conteúdo. (Spivak, 2019)

A ANPD pode expedir guia orientativo sobre bases legais de tratamento de dados biométricos em *deepfakes*. O tratamento ilegítimo de padrão facial para treinar modelo gerativo

viola arts. 7º e 11 da LGPD. A autoridade poderá aplicar multa de até 2% do faturamento do grupo econômico. Tal sanção potencializa efeito dissuasório. A integração de tutela de dados e personalidade reforça coerência sistêmica. Evita-se sobreposição normativa.

Medidas de reparação não pecuniária incluem direito de resposta audiovisual, com veiculação em mesma plataforma e alcance equivalente. O Judiciário pode determinar algoritmos de impulsionamento para maximizar correção da narrativa. Embora complexo, o STJ já admitiu esse remédio em casos de *fake news* políticas. (Mulholland; Oliveira, 2021) O *deepfake*, por sua natureza audiovisual, exige resposta na mesma linguagem. Esse paradigma assegura paridade de armas comunicacional. Preserva efetividade do direito de resposta.

A arbitragem e a mediação on-line podem ser canais céleres de resolução. Cláusulas *medi-arb* em termos de uso das plataformas poderiam prever câmaras especializadas em *deepfakes*. A decisão arbitral seria exequível internacionalmente via Convenção de Nova Iorque. Esse expediente evita morosidade estatal e permite expertise técnica no painel arbitral. Entretanto, deve-se garantir equilíbrio entre partes, pois vítima costuma ser hipossuficiente. O controle judicial mínimo assegura devido processo.

Seguradoras estudam produtos de "imagem *cyber-risk*" que cobrem custos de remoção de *deepfakes*. A precificação depende de modelos atuariais que considerem profissão, exposição midiática e histórico de ataques. A disseminação desses seguros pode transferir parte do risco econômico, mas não exime responsáveis diretos. Regulador deve estabelecer critérios para evitar seleção adversa. Incentivos fiscais poderiam estimular contratação por figuras públicas. Essa inovação financeira complementa tutela civil.

Regulamentações estaduais norte-americanas, como a emenda de Minnesota, nos Estados Unidos da América, mostram que criminalização eleitoral específica reduz *window dressing* político. (Minnesota, 2023) Estudos empíricos indicam queda de 15% na difusão de *deepfakes* eleitorais após vigência da lei. A replicação dessa política no Brasil pode ser calibrada a intervalos eleitorais definidos pelo TSE. Cria-se cultura de intolerância jurídica a manipulação audiovisual. As plataformas tenderão a banir proativamente conteúdos antes do prazo fatal. A estatística favorece adoção de norma análoga.

A doutrina vem defendendo escalonamento de responsabilidade conforme grau de controle do agente sobre a informação. O criador detém controle total, o provedor de IA parcial e a plataforma residual. Essa matriz permite gradação de culpa e quantificação diferenciada de danos regressivos. Reforça-se justiça distributiva na alocação de custos sociais. O sistema incentiva prevenção no ponto de maior eficiência. Promove análise econômica do direito.

O princípio da cooperação inter-reguladores exige diálogo entre ANPD, ANATEL, Ministério da Justiça e TSE. Cada ente detém competência parcial sobre o fenômeno. A coordenação evita lacunas e sobreposição de sanções. Sugere-se criação de comitê interagências para uniformizar protocolos de remoção e partilha de informações. O poder moderador desse comitê mitigaria conflitos de atribuição. Fortalece governança multissetorial.

A dimensão internacional do *deepfake* impõe cooperação jurídica entre Estados. A Convenção de Budapeste sobre Cibercrime, internalizada e promulgada no Brasil pelo Decreto nº 11.491, de 12 de abril de 2023, já fornece base para pedidos de preservação de dados e assistência mútua. (Brasil, 2023) A participação do Brasil na Segunda Convenção Adicional facilitaria rastreio de criadores situados no exterior. Tratados de extradição devem contemplar delito de falsificação digital. A harmonização tipológica agiliza persecução transfronteiriça. Reduz-se impunidade.

Plataformas podem utilizar inteligência artificial inversa para detectar e sinalizar *deepfakes* recém-gerados. Técnicas de *fingerprinting* e análise de inconsistências de iluminação alcançam taxa de acerto de 96% em *benchmarks*. Implantar esses sistemas como pré-processamento de upload reduz exposição inicial ao dano. O custo computacional justifica-se pela mitigação de responsabilidades futuras. Incentivos regulatórios, como redução de multas, podem acelerar adoção. Sinergia tecnológica e jurídica fortalece ecossistema de confiança.

A educação midiática integra estratégia de longo prazo. Políticas públicas devem incluir alfabetização crítica em tecnologia e verificação de fontes desde o ensino fundamental. Consumidores mais conscientes reduzem circulação de conteúdo falso. Essa dimensão preventiva dialoga com função social do ensino. O investimento governamental em campanhas de conscientização digital complementa repressão jurídica. Constrói-se resiliência social.

Finalmente, o regime de responsabilidade proposto conjuga obrigações de resultado para criadores e deveres de meio para plataformas, modulados segundo capacidade técnica. (Renaud, 2019) A equivalência de esforços evita descabida imputação de risco integral a atores de poder moderado. Adota-se análise contextual de razoabilidade. O Judiciário desempenha papel de calibragem pontual. A doutrina deverá consolidar critérios uniformes. O sistema ganha previsibilidade. (Geng, 2023)

Conclui-se que a corregulação, aliada a instrumentos civis e administrativos, oferece desenho institucional adequado ao desafio dos *deepfakes*. A integração de rotulagem obrigatória, resposta rápida e responsabilidade solidária compõe mosaico normativo proporcional. O capítulo seguinte sintetiza achados e apresenta recomendações.

4. CONCLUSÃO

O estudo confirmou a hipótese de insuficiência parcial do arcabouço jurídico brasileiro frente aos *deepfakes*. Observou-se que o atual modelo de notificação judicial é lento e favorece perpetuação do dano. A adoção de rotulagem obrigatória, já prevista no AI Act europeu, mostra-se caminho inevitável. Igualmente, a inversão dinâmica do ônus probatório equilibra assimetrias tecnológicas. A responsabilidade objetiva reforça função preventiva da tutela civil. O sistema jurídico deve evoluir sem sacrificar garantias constitucionais.

Recomenda-se atualização do PL 2630/2020 para inserir disciplina específica de conteúdo sintético. Devem-se prever obrigações de rastreabilidade, janelas eleitorais e canais prioritários de denúncia. A ANPD deveria editar guia sobre tratamento de dados biométricos em *deepfakes*. No âmbito processual, cabe ao CNJ incentivar varas cíveis a adotar protocolos de urgência para remoção global. Essas medidas podem reduzir litigiosidade e assegurar efetividade *ex ante*. Representam resposta coordenada.

Propõe-se criação de tipo penal autônomo para produção e difusão de *deepfake* lesivo, com pena proporcional e foco em conduta dolosa. Contudo, a ênfase permanece na responsabilização civil e administrativa. O direito criminal atua subsidiariamente, evitando hipertrofia. A coexistência de esferas sancionatórias amplia dissuasão sem redundância excessiva. A harmonização respeita princípio da intervenção mínima. Garante-se proporcionalidade.

O artigo delineou matriz de responsabilização escalonada entre criadores, provedores de IA, plataformas e facilitadores de pagamento. Essa arquitetura distribui custos conforme eficiência preventiva. Instrumentos de *bug bounty*, seguros *cyber-risk* e educação midiática complementam abordagem jurídica. A cooperação interagências e internacional integra a solução. Trata-se de desafio multifacetado que requer sinergia normativa, tecnológica e social. O modelo proposto oferece roteiro plausível.

Em síntese final, a tutela da honra e da imagem na era dos *deepfakes* exige evolução dogmática aliada a inovação regulatória. O direito civil, ancorado em princípios constitucionais, mantém aptidão para proteção efetiva, desde que atualizado à escala algorítmica. A conjugação de prevenção e reparação céleres resguarda dignidade humana e confiança social. Ao mesmo tempo, preserva liberdade de expressão responsável e incentiva inovação ética. O futuro da proteção da personalidade depende da coragem institucional de adotar tais reformas. Eis o imperativo normativo deste século digital.

REFERÊNCIAS

BRASIL. *Constituição (1988).* Constituição da República Federativa do Brasil. Diário Oficial da União, Brasília, DF, 5 out. 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 5 jul. 2025.

BRASIL. *Decreto nº 11.491, de 12 de abril de 2023*. Promulga a Convenção sobre o Crime Cibernético, firmada pela República Federativa do Brasil, em Budapeste, em 23 de novembro de 2001. Diário Oficial da União: Seção 1, Brasília, DF, 13 abr. 2023. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2023/decreto/d11491.htm. Acesso em: 5 jul. 2025.

BRASIL. *Decreto-Lei n. 2.848, de 7 de dezembro de 1940*. Código Penal. Diário Oficial da União, Brasília, DF, 31 dez. 1940. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto-lei/del2848.htm. Acesso em: 5 jul. 2025.

BRASIL. *Lei n. 10.406, de 10 de janeiro de 2002.* Código Civil. Diário Oficial da União, Brasília, DF, 11 jan. 2002. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/2002/l10406compilada.htm. Acesso em: 5 jul. 2025.

BRASIL. *Lei n. 12.965, de 23 de abril de 2014*. Marco Civil da Internet. Diário Oficial da União, Brasília, DF, 24 abr. 2014. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Acesso em: 5 jul. 2025.

BRASIL. *Lei n. 13.709, de 14 de agosto de 2018.* Lei Geral de Proteção de Dados Pessoais (LGPD). Diário Oficial da União, Brasília, DF, 15 ago. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 5 jul. 2025.

BRASIL. *Projeto de Lei n. 2.630, de 2020 (PL das Fake News).* Dispõe sobre a Liberdade, Responsabilidade e Transparência na Internet. Senado Federal, Brasília, DF, 2020. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=8110634. Acesso em: 5 jul. 2025.

BRASIL. Supremo Tribunal Federal. *Recurso Extraordinário n. 1.037.396/SC (Tema 987 da Repercussão Geral).* Rel. Min. Dias Toffoli, j. 26 jun. 2025. Disponível em: https://noticias.stf.jus.br/postsnoticias/stf-define-parametros-para-responsabilizacao-de-plataformas-por-conteudos-de-terceiros/. Acesso em: 5 jul. 2025.

CONSELHO DA EUROPA. *Convenção sobre o Crime Cibernético (Budapeste)*, 23 nov. 2001. Tratado ETS n. 185. Budapeste, 2001. Disponível em: https://rm.coe.int/1680081561. Acesso em: 5 jul. 2025.

DE RUITER, Adrienne. The Distinct Wrong of Deepfakes. *Philosophy & Technology*, v. 34, 2021, p. 1311-1332.

GENG, Yinuo. Comparing "Deepfake" Regulatory Regimes in the United States, the European Union and China. *Georgetown Law Technology Review*, v. 7, 2023, p. 157-177.

MEDON, Filipe. O direito à imagem na era das deepfakes. *Revista Brasileira de Direito Civil*, Belo Horizonte, v. 27, p. 251-277, jan./mar. 2021.

MINNESOTA (EUA). *Minnesota Statute 609.771 (HF 1370/2023)*. Lei que cria infrações eleitorais envolvendo deepfakes. St. Paul, 26 mai 2023. Disponível em: https://www.revisor.mn.gov/bills/bill.php?b=house&f=hf1370&ssn=0&y=2023. Acesso em: 5 jul. 2025.

MULHOLLAND, Caitlin; OLIVEIRA, Samuel Rodrigues de. Uma nova cara para a política? Considerações sobre deepfakes e democracia. *Revista Direito e Política*, Brasília, v. 18, n. 99, jul./set. 2021, p. 368-396.

PINTO, Felipe Chiarello de Souza; OLIVEIRA, Gabriela Franklin de. Não acredite em tudo que vê: Deepfake pornography e responsabilidade civil no ordenamento jurídico brasileiro. *Direito e Política*, v. 18, n. 2, 2023, p. 427-449.

RENAUD, Lauren. Will You Believe It When You See It? How and Why the Press Should Prepare for Deepfakes. *Georgetown Law Technology Review*, v. 4, n. 1, 2019, p. 241-261.

SILVEIRA, Pedro Henrique Scoralick; FALEIROS JÚNIOR, José Luiz de Moura. Inteligência artificial e a exploração do direito de imagem a partir dos 'deepfakes'. *Inova Jur - Revista Jurídica da UEMG*, Belo Horizonte, v. 3, n. 1, p. 1-19, jan./jun. 2024.

SILVEIRA, Pedro Henrique Scoralick; AUTO, Ana Luíza Alves Ferreira Silva; FALEIROS JÚNIOR, José Luiz de Moura. Violações a direitos da personalidade oriundas de 'deep fake porn'. In: *V Congresso Internacional de Direito e Inteligência Artificial*, 2024, Belo Horizonte/MG, Brasil. JUNQUILHO, Tainá Aguiar; SILVA, Paula Guedes Fernandes da; RIBEIRO, Fernanda (Org.). Regulação da inteligência artificial - I [recurso eletrônico]. Belo Horizonte: Skema Business School, 2024. v. 7. p. 23-29.

SPIVAK, Russell. "Deepfakes": The Newest Way to Commit One of the Oldest Crimes. *Georgetown Law Technology Review*, v. 3, n. 2, 2019, p. 339-400.

UNIÃO EUROPEIA. *Regulamento (UE) 2022/2065, de 19 out. 2022.* Digital Services Act. Jornal Oficial da União Europeia, Luxemburgo, L 277, 27 out. 2022. Disponível em: https://eur-lex.europa.eu/eli/reg/2022/2065/oj. Acesso em: 5 jul. 2025.

UNIÃO EUROPEIA. *Regulamento (UE) 2024/1689, de 13 jun. 2024*. Artificial Intelligence Act. Jornal Oficial da União Europeia, Luxemburgo, L 2024/1689, 12 jul. 2024. Disponível em: https://eur-lex.europa.eu/eli/reg/2024/1689/oj. Acesso em: 5 jul. 2025.

UNIÃO EUROPEIA. Comissão Europeia. *Strengthened Code of Practice on Disinformation*. Bruxelas, 2022. Disponível em: https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation. Acesso em: 5 jul. 2025.

UNITED STATES OF AMERICA. Congress. *S. 560 – SAFE TECH Act.* 118th Congress, 1st Session, 28 feb. 2023. Washington, DC, 2023. Disponível em: https://www.congress.gov/bill/118th-congress/senate-bill/560/text. Acesso em: 5 jul. 2025.

UNITED STATES OF AMERICA. Congress. *S. 146 – TAKE IT DOWN Act.* 119th Congress, 16 jan. 2025. Washington, DC, 2025. Disponível em: https://www.congress.gov/bill/119th-congress/senate-bill/146. Acesso em: 5 jul. 2025.

ZIOBROŃ, Agata. Political deepfake. Remarks de lege lata and postulates de lege ferenda. *Studia Prawnicze. Rozprawy e Materiais*, Cracóvia, n. 1 (34), 2024, p. 80-95.