DEEPFAKE COM GANs: UMA PERSPECTIVA JURÍDICA PARA A RESPONSABILIZAÇÃO PELOS DEEPFAKES

LARISSA DA SILVA MENDES[1]; LUIS EDUARDO RASCH[2]; MANOELA VIERA DE MOURA[3]; MICAEL FONTOURA MENDES[4]; MURILO COSTA SALEM[5]; ANDERSON PRIEBE FERRUGEM[6]

[1]*Universidade Federal de Pelotas – larissa.mendes@ufpel.edu.br*
[2]*Universidade Federal de Pelotas – lerasch@inf.ufpel.edu.br*
[3]*Universidade Federal de Pelotas – mvmoura@inf.ufpel.edu.br*
[4]*Universidade Federal de Pelotas – micael.mendes@ufpel.edu.br*
[5]*Universidade Federal de Pelotas – mcsalem@inf.ufpel.edu.br*
[6]*Universidade Federal de Pelotas – ferrugem@inf.ufpel.edu.br*

1. INTRODUÇÃO

O rápido avanço das Redes Generativas Adversariais (Generative Adversarial Networks - GANs), um subcampo da inteligência artificial, possibilitou a criação de mídias sintéticas altamente realistas, conhecidas como *deepfakes*. Embora esta tecnologia ofereça aplicações inovadoras, ela também apresenta desafios legais e éticos significativos, incluindo desinformação, violações de privacidade e fraude de identidade (KHIMI et al., 2024; STORY; JENKINS, 2023; TRELEAVEN et al., 2023).

Este artigo examina as implicações jurídicas dos *deepfakes*, com foco nos arcabouços regulatórios existentes e propondo medidas para mitigar seu uso indevido. Analisamos métodos de detecção, esforços legislativos no Brasil e a necessidade de uma abordagem equilibrada que combine soluções técnicas, responsabilização legal e conscientização pública. Nossas conclusões destacam a urgência de regulamentações atualizadas para proteger os direitos individuais e a integridade democrática na era do conteúdo gerado por IA.

2. METODOLOGIA

O trabalho foi desenvolvido a partir de uma pesquisa interdisciplinar, com ênfase nas áreas da Ciência da Computação e do Direito. Desse modo, a metodologia adotada consistiu na análise bibliográfica e documental, abrangendo artigos científicos nacionais e internacionais, livros especializados, matérias jornalísticas e dispositivos legais relacionados ao uso de inteligência artificial na geração de mídias sintéticas.

Assim, a produção do artigo exigiu constante comunicação entre os co-autores, permitindo a integração entre os conhecimentos técnicos e jurídicos e a construção de um estudo aprofundado e crítico sobre o tema.

3. RESULTADOS E DISCUSSÃO

A partir da pesquisa, foi possível identificar que o avanço das tecnologias generativas, especialmente as GANs, tem ampliado significativamente o uso de

deepfakes em contextos sociais sensíveis, como política, pornografia não consensual e fraudes financeiras (KHIMI et al., 2024; STORY; JENKINS, 2023; TRELEAVEN et al., 2023). Desse modo, esses conteúdos, cada vez mais realistas, dificultam a identificação de falsificações, comprometendo a confiança pública e gerando graves impactos éticos e jurídicos (TUYSUZ; KILIÇ, 2023).

Assim, foram discutidos casos emblemáticos, como o vídeo manipulado atribuído ao ministro Fernando Haddad (LEITÃO, 2024) e fraudes financeiras com uso de voz e imagem artificialmente clonadas (CNN, 2024). Dessa forma, a análise revelou lacunas regulatórias na legislação brasileira, que ainda não dispõe de dispositivos específicos para punir ou prevenir a produção e disseminação de conteúdo sintético danoso (AFFONSO, 2021; FERREIRA; LEME, 2023; MARANHÃO et al., 2021).

Nesse contexto, o trabalho destaca a importância de combinar mecanismos legais, como a criminalização específica de *deepfakes* e a responsabilização de plataformas, com medidas técnicas, como marcações digitais e ferramentas de detecção automática (ZHANG et al., 2019; GUO et al., 2022; GUO; LI; LYU, 2021). A discussão evidencia, assim, a urgência de um ecossistema regulatório capaz de enfrentar os desafios trazidos pelas mídias sintéticas (FLORIDI et al., 2018; DE MATTOS et al., 2024).

4. CONCLUSÕES

A disseminação de *deepfakes* desafia os pilares jurídicos da democracia, da verdade e da dignidade humana (LEME e FERREIRA, 2023). Dessa forma, conclui-se que, apesar da existência de proteção constitucional e civil ao direito à imagem, a velocidade de propagação e a sofisticação técnica dos conteúdos artificiais exigem regulações específicas, atualizadas e eficazes.

Enfim, a legislação em tramitação representa um avanço importante, especialmente ao prever crimes autônomos para *deepfakes* e responsabilizar os intermediários digitais. No entanto, é preciso articular isso a um ecossistema de governança da IA que envolva não apenas sanções, mas também educação digital, transparência algorítmica e participação social. Assim, como concluem os autores analisados, combater os efeitos nocivos das *deepfakes* é essencial não apenas para proteger a imagem individual, mas para preservar a própria legitimidade das instituições democráticas frente à era da inteligência artificial.

5. REFERÊNCIAS BIBLIOGRÁFICAS

AFFONSO, F. J. M. O direito à imagem na era das deep fakes. *Revista Brasileira de Direito Civil*, São Paulo, v.27, n.1, p.251-251, 2021.

ARJOVSKY, M.; BOTTOU, L. Towards principled methods for training generative adversarial networks. *arXiv preprint*, arXiv:1701.04862, 2017.

ARJOVSKY, M.; CHINTALA, S.; BOTTOU, L. Wasserstein GAN. *arXiv preprint*, arXiv:1701.07875, 2017.

BENDER, E. M. et al. On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, New York, v.?, n.?, p.610–623, 2021.

BOMMASANI, R. et al. On the opportunities and risks of foundation models. *arXiv preprint*, arXiv:2108.07258, 2021.

BROWN, T. B. et al. Language models are few-shot learners. In: *Advances in Neural Information Processing Systems*, v.33, p.1877–1901, 2020.

BRUNDAGE, M. et al. Toward trustworthy AI development: Mechanisms for supporting verifiable claims. *arXiv preprint*, arXiv:2004.07213, 2022.

CARLINI, N. et al. Extracting training data from diffusion models. *arXiv preprint*, arXiv:2301.13188, 2023.

CHEN, T. C.; LIU, M. Y. et al. Semantic image synthesis with spatially-adaptive normalization. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, v.?, n.?, p.2337–2346, 2018.

CNN. Deepfake video call scam tricks Hong Kong employee into paying out $25 million. *CNN International*, 2024. Disponível em: https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk. Acesso em: 23 jul. 2025.

CRAWFORD, K. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press, 2021.

DE MATTOS, A. E. N. P.; CURTO, L. V.; MUSSALLAM, M. S. Inteligência artificial e o direito digital. *Revista Políticas Públicas & Cidades*, v.13, n.2, p.1–36, 2024.

DEEPFAKE Pornography Report. Deepfake laws risk creating more problems than they solve. *RTP / Federalist Society*, 2024. Disponível em: https://rtp.fedsoc.org/wp-content/uploads/Paper-Deepfake-Laws-Risk-Creating-More-Problems-Than-They-Solve.pdf. Acesso em: 23 jul. 2025.

GOIÁS (Estado). Lei Complementar nº 205/2025. *Legisla Goiás*, 2025.

CÂMARA DOS DEPUTADOS. Portal da Câmara dos Deputados. *Brasília*, 2025.

FAGUNDES, E.; LUPA. Sites fraudulentos voltam a usar advogados para se passar pelos correios. 2025.

FERREIRA, A. P.; LEME, C. da S. O fenômeno da deep fake no contexto eleitoral e seus efeitos no estado democrático de direito. *Boletim IBCCRIM*, São Paulo, v.31, n.363, p.21–23, 2023.

FLORIDI, L. et al. AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, Dordrecht, v.28, n.4, p.689–707, 2018.

GENERATED PHOTOS. This person does not exist. 2019. Disponível em: https://thispersondoesnotexist.com. Acesso em: 24 jul. 2025.

GOODFELLOW, I. et al. Generative adversarial nets. In: *Advances in Neural Information Processing Systems*, p.2672–2680, 2014.

GUO, H.; LI, Y.; LYU, S. Eyes tell all: Irregular pupil shapes reveal GAN-generated faces. *arXiv preprint*, arXiv:2109.00162, 2021.

JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. *Science*, New York, v.349, n.6245, p.255–260, 2015.

KARRAS, T.; LAINE, S.; AILA, T. A style-based generator architecture for generative adversarial networks. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, p.4401–4410, 2019.

KHIMI, W. et al. A systematic review on deep fake image generation, detection techniques, ethical implications, and overcoming challenges. *International Journal of Computing and Informatics*, Reino Unido, v.3, n.8, p.1–12, 2024.

LEDIG, C. et al. Photo-realistic single image super-resolution using a generative adversarial network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p.4681–4690, 2017.

LEITÃO, M. Haddad é a primeira vítima da nova política de Zuckerberg. *Veja*, São Paulo, 2024. Disponível em: https://veja.abril.com.br/coluna/matheus-leitao/haddad-e-a-primeira-vitima-da-nova-politica-de-zuckerberg/. Acesso em: 23 jul. 2025.

MARANHÃO, J. S. A.; FLORÊNCIO, J. A.; ALMADA, M. Inteligência artificial aplicada ao direito e o direito da inteligência artificial. *Suprema: Revista de Estudos Constitucionais*, v.1, p.154–180, 2021.

RAMESH, A. et al. Zero-shot text-to-image generation. *arXiv preprint*, arXiv:2102.12092, 2021.

REUTERS; G1. Facebook admite erro em atraso para sinalizar vídeo falso nos EUA. 2019.

AGÊNCIA SENADO. Projetos buscam restringir manipulação de imagens com inteligência artificial. 2024.

STORY, D.; JENKINS, R. Deepfake pornography and the ethics of non-veridical representations. *Philosophy & Technology*, Berlim, v.36, p.1–22, 2023.

TRELEAVEN, P. et al. The future of cybercrime: AI and emerging technologies are creating a cybercrime tsunami. *SSRN*, 2023.

TUYSUZ, M. K.; KILIÇ, A. Analyzing the legal and ethical considerations of deepfake technology. *Interdisciplinary Studies in Society, Law, and Politics*, v.2, n.2, p.4–10, 2023.

ZHANG, R.; ZHANG, J.; ZHANG, D. Invisible steganography via generative adversarial networks. *Neurocomputing*, Amsterdã, v.332, p.125–134, 2019.