IRIE
International Review of Information Ethics Vol. 33 (03/2024)

Author(s): Alana Maria Passos Barreto

Para além das *fake news*: breves apontamentos sobre a inteligência artificial imitativa

Beyond *fake news*: brief notes on imitative artificial intelligence

Resumo:

O rápido aumento da inteligência artificial (IA) criou muitas oportunidades em toda a sociedade. No entanto, essas rápidas mudanças também levantam profundas preocupações comportamentais. De tal maneira, em vista da necessidade do Direito se adaptar aos novos fatos que surgem dentro da sociedade, o presente artigo se propõe a analisar a regulação sobre a aplicação da IA em imagens, vídeos e áudios *deepfakes*. Utilizou-se a pesquisa qualitativa, de caráter descritivo, através de levantamento bibliográfico e documental para tratar do estado da arte das *deepfakes*. A inteligência artificial tem sido amplamente utilizada, mas a ausência de regulamentação adequada pode resultar em violações de direitos alheios e o uso indevido dessa tecnologia, pode causar significativos em danos.

Palavras-chave: *Deepfakes*; Informação; Inteligência Artificial.

Abstract:

The rapid rise of artificial intelligence (AI) has created many opportunities across society. However, these rapid changes also raise deep behavioral concerns. In this way, in view of the need for Law to adapt to the new facts that arise within society, this paper proposes to analyze the regulation on the application of AI in deepfakes images, videos and audios. Qualitative, descriptive research was used, through bibliographic and documentary research to address the state of the art of *deepfakes*. The artificial intelligence has been widely used, but the absence of adequate regulation can result in violations of the rights of others and the misuse of this technology can cause significant damage.

Keywords: *Deepfakes*; Information; Artificial intelligence.

Agenda:

Author(s):

ISSN 1614-1687

Considerações iniciais

O rápido aumento da inteligência artificial (IA) criou muitas oportunidades em toda a sociedade. No entanto, essas rápidas mudanças também levantam profundas preocupações comportamentais. Isso decorre do potencial que os sistemas de IA têm para incorporar vieses e de como isso impacta nos direitos humanos. Dessa forma, os riscos associados à IA já estão no plano central do debate.

Em março de 2023, especialistas e pesquisadores de inteligência artificial assinaram uma carta aberta pedindo a todos os laboratórios de inteligência artificial que suspendam pelo prazo mínimo de seis meses o desenvolvimento de sistemas de IA, como o popular *ChatGPT*, da OpenAI. De tal maneira, em vista da necessidade do Direito se adaptar aos novos fatos que surgem dentro da sociedade, o presente artigo se propõe a analisar a regulação sobre a aplicação da IA em imagens, vídeos e áudios *deepfakes*.

E para realizar essa análise, o texto a seguir divide-se em dois momentos, além das considerações iniciais e finais, sendo a primeira parte destinada a compreender o tratamento dispensado à reconstrução digital de imagens a partir de tecnologias de inteligência artificial. Já a segunda parte é destinada a contextualizar a disciplina legal do Direito Digital brasileiro, em razão da dificuldade de encaixar as *deepfakes* nas legislações existentes.

Para fins metodológicos, este artigo de revisão, utilizou a pesquisa qualitativa com caráter descritivo para tratar do estado da arte das *deepfakes*. De tal maneira, para a análise dos objetivos, a coleta de dados foi essencial através do levantamento bibliográfico e documental. Assim, a metodologia adotada neste trabalho não é apenas uma mera descrição sobre os dados das fontes pesquisadas, de modo que fora estabelecido relações e comparações entre as informações reunidas.

Dessa forma, a IA tem sido amplamente utilizada para simular a aparência de pessoas públicas, famosas e, até mesmo, falecidas, criando debates em diversas áreas do cotidiano. No entanto, a ausência de regulamentação adequada pode resultar em violações de direitos alheios e o uso indevido dessa tecnologia, pode causar significativos em danos.

Deepfakes: a inteligência artificial aplicada à desinformação

Em 2018, o *BuzzFeed* publicou uma matéria intitulada *A Belgian Political Party Is Circulating A Trump Deep Fake Vídeo* (Um partido político belga está circulando um vídeo Deepfake de Trump, tradução nossa) e criou um viral diante do termo *deepfake*. Ao ler a manchete, espera-se que trate de uma campanha de propaganda política de alta tecnologia, utilizando inteligência artificial para colocar palavras na boca do ex-presidente dos Estados Unidos, Donald Trump, e enganar os eleitores, mas ao assistir o vídeo fica evidente o tom humorístico, com representação vocal exagerada e efeitos de computador irrealistas (Lytvynenko, 2018).

Originariamente, o termo *deepfake* surgiu de um usuário do *Reddit*, em 2017, que usou ferramentas de IA disponíveis à época do mercado para colar rostos de celebridades em videoclipes pornográficos. Embora, o termo tenha sido aplicado a falsificações pornográficas, foi adotado como abreviação para uma ampla variedade de vídeos e imagens editadas. Atualmente, as *deepfakes* incluem trocas de rosto, cópias a voz de alguém, reencenação facial – ou seja, mapear o rosto de alguém e manipulá-lo –, e sincronização – caracterizado pelo vídeo criado de alguém falando a partir de áudio e imagens de seu rosto.

O termo *deepfake* é uma junção de *deep learning* e *fake* e refere-se a multimídia gerada por IA (imagens, vídeos, áudios e textos) que são potencialmente enganosos (Vincent, 2018), sua maior característica é o processo de edição automatizada usando técnicas de Inteligência Artificial (IA), sob o manto do aprendizado profundo. O que inicialmente beira o humor viral, logo se torna um mecanismo criminoso, de modo que a tecnologia *deepfake* pode gerar, por exemplo, um vídeo pornográfico ou político de alguém dizendo qualquer coisa, sem o consentimento da pessoa cuja imagem e voz estão envolvidas. Muito embora os primeiros exemplos sejam de figuras públicas com seus rostos entrelaçados em vídeos humorísticos, a tecnologia pode

ser utilizada para pornografia de vingança, bullying, provas falsas em processos judiciais, sabotagem política, propaganda terrorista, entre outros (MARAS; ALEXANDROU, 2019).

Patrini (2018) acredita que um *deepfake* deve incluir algum componente automatizado e aprendido. Por sua vez, Brundage (2018) coloca que o termo *deepfake* não tem limites distintos, mas geralmente se refere a um subconjunto de vídeos falsos que se aproveitam de um aprendizado profundo para tornar o processo de falsificação mais fácil.

Compreende-se que, vídeos e imagens editados com softwares existentes, como *Adobe Photoshop* e *After Effects*, não são *deepfakes*. No entanto, pela definição de Patrini (2018), essa regra não é confiável, tendo em vista que esses programas automatizaram parte do processo de edição e, brevemente oferecerão recursos baseados em IA, como o *Snapchat* que usa técnicas de IA para aplicar filtros nos rostos das pessoas, porém não são chamados de *deepfakes*. O mesmo processo acontece com os *animojis* da *Apple*, que poderiam ser considerados desenhos animados *deepfakes*.

Há diversas maneiras de criar uma *deepfake*, e uma das mais adotadas faz uso de Redes Geradoras Adversariais (GAN), que é composta por duas redes neurais profundas, uma chamada de gerador e a outra de discriminador. A primeira gera os dados, enquanto a segunda treina o uso de dados, buscando a verdade. A *deepfake* só estará apta a produzir seus efeitos quando o discriminador não conseguir mais distinguir se o conteúdo é verdadeiro ou falso.

Ao entender o uso da IA em aplicativos de edição ou entretenimento, nota-se que ao tratar de *deepfakes* considera-se o conteúdo com potencial (malicioso) de enganar alguém e até mesmo de afetar significativamente suas vidas. Como também pode ser levado a níveis coletivos e difusos ainda maiores, pois pode influenciar a opinião pública, causando inseguranças políticas, ou sendo usado em um tribunal como evidência falsa.

No caso das falsificações pornográficas, afeta a vida íntima, a privacidade e a honra do indivíduo. O que se extrai dessas definições é que a *deepfake* é composta pela IA que desenvolve um procedimento automatizado e potencialmente enganoso a terceiros.

Os algoritmos existentes usados para criação de *deepfakes* podem ser divididos em duas categorias: troca de rosto e reconstituição de rosto. Tratando inicialmente sobre a troca de rostos, o *Faceswap-GAN* (2018), uma versão melhorada dos algoritmos *deepfake* originais, e foi proposto em 2018, e posteriormente o *VGGFace* (2019) foi implementado para gerar faces mais realistas.

Por sua vez, o *DeepFaceLab* é uma estrutura de geração de *deepfake* de código aberto e foi projetada para fornecer uma plataforma interativa de fácil manuseio para pessoas sem conhecimento profissional (PETROV et al. 2020). Já o *FaceShifter* foi proposto para gerar rostos trocados de alta fidelidade ao realizar uma integração abrangente de atributos faciais (Lingzhi *et a*l., 2019).

Os experimentos dessas *deepfakes* que utilizam as IA demonstram avanços superiores em comparação com algoritmos de troca de face existentes. Os vídeos gerados por abordagens *deepfake* recentes têm sido extremamente realistas, dificilmente distinguidos pelos olhos humanos. Por outro lado, os algoritmos de reconstituição facial tentam controlar as expressões das pessoas nos vídeos, o que significa que os invasores podem gerar vídeos manipulando alguém para fazer algo que não existe (Peipeng; Zhihua; Jianwei & Yujiang, 2021).

FakeApp e *FaceSwap* são exemplos de aplicativos baseados em *deepfake*, e similares continuam a aparecer – em 2019, surgiu um aplicativo para despir roupas de imagens chamado *Deepnude*. Além de prejudicar a privacidade pessoal, os vídeos gerados por esses aplicativos podem interferir em campanhas políticas e na opinião pública. Desse modo, a detecção de conteúdo *deepfake* tornou-se um dos principais problemas para indivíduos, empresas e governos em todo o mundo (Peipeng; Zhihua; Jianwei & Yujiang, 2021).

Os primeiros algoritmos de reconstituição facial se utilizaram do modelo facial, que foi modificado sob diferentes parâmetros de expressão, ou seja, um modelo paramétrico é aproveitado para ajustar imagens faciais. Esses métodos podem gerar imagens faciais com alto realismo, mas os resultados obtidos muitas vezes carecem de coerência temporal. Nos últimos anos, a pesquisa sobre reconstituição facial foi desenvolvida à medida que a capacidade de computação aumentou (Peipeng; Zhihua; Jianwei & Yujiang, 2021).

No entanto, para realizar a reconstituição facial monocular em tempo real foi proposto o *Face2Face*, em que trouxe uma nova abordagem de agrupamento baseada em modelos não rígidos globais para reconstruir as características faciais dos atores alvo e fonte (THIES et al., 2016). Ao mesmo tempo em que é realizada uma técnica de transferência de deformação subespacial projetada para realizar a transferência de expressão entre os atores fonte e alvo. Nota-se que, o *Face2Face* já alcançou um desempenho bastante notável (Peipeng; Zhihua; Jianwei & Yujiang, 2021).

Um breve retrospecto sobre a legislação digital brasileira

Os perigos da conectividade são extremamente subestimados, isso porque a desatenção tardia acarretou uma grande quantidade de vítimas de furto, da perda de dados privados, de informações falsas. No entanto, a sociedade de informação inserida numa típica sociedade de risco (Beck, 2011) permanece ignorante sobre sua vulnerabilidade.

Schawb (2016) alerta que a dinâmica da partilha das mídias sociais pode enviesar a tomada de decisões e causar riscos para sociedade civil, pois a grande quantidade de conteúdo disponível nos canais digitais é capaz de polarizar as fontes de informação do indivíduo. Nesse viés, a relação entre as plataformas e a disseminação das informações convergem de modo que a combinação de ausência de regras e algoritmos encorajam a rápida difusão de conteúdos de qualidade questionável.

Dessa forma, é importante compreender que não é adequado empregar argumentos jurídico-técnicos para 'remendar' uma base legal 'implícita', dado que o poder, a escala e a intromissão dessas tecnologias criam sérias ameaças aos direitos e liberdades democráticas (Grohman, 2020). "A falta de parâmetros legais deixa em aberto uma lacuna jurídica, regulatória e ética, com as más consequências que o uso de sistemas de IA sem governança pode trazer" (Transparência Brasil, 2020, p. 5).

A IA possui potencial para reduzir o custo de ataques, o que facilita a automatização das tarefas que exigiam trabalho humano, logo, é necessário que (i) os pesquisadores de IA devem reconhecer como seu trabalho pode ser usado de forma maliciosa; (ii) os formuladores de políticas precisam aprender com especialistas técnicos sobre essas ameaças; (iii) o mundo da IA precisa aprender com os especialistas em segurança cibernética como proteger melhor seus sistemas; (iv) estruturas éticas para IA precisam ser desenvolvidas e seguidas; e (v) mais pessoas precisam estar envolvidas nessas discussões, não apenas pesquisadores e formuladores de políticas públicas, mas também especialistas em ética, empresas e a sociedade civil (BRUNDAGE, 2018).

Dessa forma, o combate às *deepfakes* é um desafio crescente em um mundo cada vez mais conectado e digital. São vários os fatores que podem interferir nesse cenário, a exemplo da (i) dificuldade na detecção, uma vez que a tecnologia de edição e manipulação de imagens e áudios está se tornando cada vez mais sofisticada e autêntica; (ii) falta de regulação específica para tratar das *deepfakes*; (iii) a dificuldade de verificação da motivação, já que muitas *deepfakes* são criadas por razões políticas ou financeiras, o que torna mais desafiador o combate a essas práticas; (iv) velocidade de propagação, que dificulta a rastreabilidade e o controle da disseminação, o que torna o combate a essas práticas ainda mais complexo.

O projeto de lei sobre a *deepfakes*

A empresa automobilística alemã, *Volkswagen*, realizou uma campanha publicitária para comemorar seu septuagésimo aniversário em que utilizou de *deepfake* para recriar a imagem da cantora Elis Regina, falecida

em 1982. No vídeo promocional, Elis é "revivida" pela IA e aparece em um dueto com sua filha Maria Rita, para ilustrar o relançamento da Kombi, e juntas interpretam a música "Como Nossos Pais", escrita por Belchior.

Em razão da ampla repercussão da campanha publicitária, o Conselho Nacional de Autorregulamentação Publicitária (Conar), uma entidade não governamental, recebeu várias reclamações de consumidores, o que levou a abertura de um processo ético (Rodrigues, 2023). As queixas levantavam questões sobre a ética do uso da IA para "dar vida" a uma pessoa falecida e até que ponto essa tecnologia pode causar confusão na percepção da realidade por parte de crianças e adolescentes.

Devido ao sucesso do comercial da *Volkswagen*, foi apresentado no Senado Federal o projeto de lei 3592/23 com o intuito de regulamentar o uso de inteligência artificial nos casos em que houver possibilidade de uso de imagem e voz de pessoa falecida. Dessa forma, o software da IA deve ter a autorização dos proprietários de qualquer imagem ou voz digitalizada que usem para fazer novos materiais, seja de pessoa viva ou não, a proposta é manter essa limitação igual existe para outras tecnologias audiovisuais. O PL ainda propõe que os herdeiros possam decidir sobre o uso da IA para representar pessoa falecida, estendendo a essa tecnologia as mesmas limitações que pesam sobre as demais.

Nesse sentido, observa-se que o projeto objetiva garantir o direito de decisão sobre o uso de imagem de si após a morte, enquanto os herdeiros serão responsáveis por cuidar do uso da imagem do falecido pela IA e terão o direito de negá-lo para usos que considerem inadequados, de modo que não havendo declaração prévia do falecido, os herdeiros poderão decidir se autorizam ou não o uso.

Considerações finais

À vista disso, a disseminação de *deepfakes* pode ter sérias consequências para a reputação e privacidade de pessoas, empresas e instituições. Apesar de ainda não possuir uma lei única, o uso da IA já é regulada no Brasil. Tendo em vista que normas como o Marco Civil da Internet, a Lei Geral de Proteção de Dados Pessoais, o Código de Defesa do Consumidor, o Código Civil e a Lei de Direitos Autorais impactam direta ou indiretamente sobre o tema. Ainda assim, existem tentativas de centralizar essa regulação em normas únicas, como o PL 21/20 – que possui um caráter principiológico – e o projeto 2338/23, baseado no relatório final da Comissão de juristas responsável por subsidiar a elaboração de substitutivo sobre IA (CJUSBIA).

No entanto, inúmeras questões ficaram de fora do breve escopo desta análise, mas merecem igual atenção. De todo modo, percebe-se que a sofisticação tecnológica demanda de respostas mais criativas, dada a velocidade na transmissão de conteúdos na internet e o potencial lesivo resultante.

Referências

BECK, U. Sociedade de risco: rumo a uma outra modernidade. São Paulo. Ed. 34. 2011.

BRASIL. Projeto de Lei n° 3592, de 2023. Brasília, DF: Senado Federal, 2023. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/158816. Acesso em: 02 ago. 2023.

BRUNDAGE, M. et al. The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. Malicious Aireport, 2018. Disponível em: https://maliciousaireport.com/. Acesso em: 15 jan. 2023.

FACESWAP-GAN. 2018. Disponível em: https://github.com/shaoanlu/faceswap-GAN. Acesso em: 27 jan. 2023.

LINGZHI, L., et al. Faceshifter: towards high fidelity and occlusion aware face swapping. Cornell University, 2019. Disponível em: https://arxiv.org/abs/1912.13457. Acesso em: 23 jan. 2023.

LYTVYNENKO, J. A Belgian Political Party Is Circulating A Trump Deep Fake Vídeo. BuzzFeed, 2018. Disponível em: https://www.buzzfeednews.com/article/janelytvynenko/a-belgian-political-party-just-published-a-deepfake-video. Acesso em: 12 jan. 2023.

MARAS, M. H.; ALEXANDROU, A. Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. International Journal of Evidence & Proof, 23(3): 255–262. 2019. https://doi.org/10.1177/1365712718807226. Acesso em: 22 jul. 2022.

PATRINI, G. Commoditisation of AI, digital forgery and the end of trust: how we can fix it. Github, 2018. https://giorgiop.github.io/posts/2018/03/17/AI-and-digital-forgery/. Acesso em: 15 jan. 2023.

PEIPENG, Y.; ZHIHUA, X.; JIANWEI, F.; YUJIANG, L. A Survey on Deepfake Video Detection. IET Biometrics, vol. 10, ed. 6, 2021, p. 607-624. Disponível em: https://doi.org/10.1049/bme2.12031. Acesso em: 27 jan. 2023.

PETROV, I., et al. DeepFaceLab: a simple, flexible and extensible face swapping framework. Cornell University, 2020. Disponível em: https://arxiv.org/abs/2005.05535. 23 jan. 2023.

RODRIGUES, E. Conar abre processo ético sobre Volkswagen ter "revivido" Elis Regina. Autopapo, Uol, 2023. Disponível em: https://autopapo.uol.com.br/curta/conar-processo-comercial-volkswagen/. Acesso em: 02 ago. 2023.

SCHWAB, K. A Quarta Revolução Industrial. Tradução Daniel Moreira Miranda. São Paulo: Edipro, 2016.

TRANSPARÊNCIA BRASIL. Recomendações de Governança: Uso da Inteligência Artificial pelo Poder Público. Transparência, 2020. Disponível em: https://www.transparencia.org.br/downloads/publicacoes/Recomendacoes_Governanca_Uso_IA_PoderPublico.pdf. Acesso em: 09 ago. 2023.

THIES, J., et al.: Face2face: real-time face capture and reenactment of RGBvideos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, p. 2387-2395, 2016.

KERAS-VGGFACE: Implementação do Vggface com Keras Framework. 2019. Disponível em: https://github.com/rmalli/keras-vggface. Acesso em: 27 jan. 2023.

VINCENT J. Why we need a better definition of 'deepfake'. The Verge, 2018. Disponível em: https://www.theverge.com/2018/5/22/17380306/deepfake-definition-ai-manipulation-fake-news. Acesso em: 10 jan. 2023.