Centro Universitário de Brasília - UniCEUB

Faculdade de Ciências Jurídicas e Sociais - FAJS
Curso de Bacharelado em Direito

MARIA FERNANDA MUGNAINI NAKANISHI

A PROBLEMÁTICA JURÍDICA DOS DEEPFAKES: UMA ANÁLISE DO USO DA INTELIGÊNCIA ARTIFICIAL NA PRODUÇÃO DE PROVAS E SUAS REPERCUSSÕES PENAIS

BRASÍLIA
2023
MARIA FERNANDA MUGNAINI NAKANISHI

A PROBLEMÁTICA JURÍDICA DOS DEEPFAKES: UMA ANÁLISE DO USO DA INTELIGÊNCIA ARTIFICIAL NA PRODUÇÃO DE PROVAS E SUAS REPERCUSSÕES PENAIS.

Artigo científico apresentado como requisito parcial para obtenção do título de Bacharel em Direito pela Faculdade de Ciências Jurídicas e Sociais - FAJS do Centro Universitário de Brasília (UniCEUB).

Orientador: Professor Dr. Victor Minervino Quintiere.

BRASÍLIA
2023

MARIA FERNANDA MUGNAINI NAKANISHI

A PROBLEMÁTICA JURÍDICA DOS DEEPFAKES: UMA ANÁLISE DO USO DA INTELIGÊNCIA ARTIFICIAL NA PRODUÇÃO DE PROVAS E SUAS REPERCUSSÕES PENAIS.

Artigo científico apresentado como requisito parcial para obtenção do título de Bacharel em Direito pela Faculdade de Ciências Jurídicas e Sociais - FAJS do Centro Universitário de Brasília (UniCEUB).

Orientador(a): Professor Dr. Victor Minervino Quintiere.

Brasília-DF, ____ de _____ de 2023.

BANCA AVALIADORA

_____

Professor Dr. Victor Minervino Quintiere.

_____

Professor(a) Avaliador(a)

A PROBLEMÁTICA JURÍDICA DOS DEEPFAKES: UMA ANÁLISE DO USO DA INTELIGÊNCIA ARTIFICIAL NA PRODUÇÃO DE PROVAS E SUAS REPERCUSSÕES PENAIS.

Maria Fernanda Mugnaini Nakanishi

Resumo: Trata-se de artigo apresentado como requisito para conclusão do Curso de Direito da Faculdade de Ciências Jurídicas e Sociais do Centro Universitário de Brasília. Esta pesquisa teve como objetivo apresentar a problemática jurídica dos deepfakes fazendo uma análise da questão do uso da Inteligência Artificial na produção de provas e suas repercussões no âmbito jurídico. Esse estudo, portanto, teve como problema de pesquisa esclarecer a extensão em que o fenômeno dos *deepfakes* consegue influenciar negativamente na produção de provas no sistema penal brasileiro. A presente pesquisa bibliográfica foi baseada numa abordagem qualitativa e estruturada em três capítulos: o primeiro capítulo teve como objetivo elucidar a respeito dos conceitos, classificação e regras do papel das provas no processo penal. O segundo, por sua vez, explicou sobre a relação do direito penal e da tecnologia, fazendo uma análise dos novos delitos digitais e as leis e normas que regem essa relação. Por último, abordou-se sobre os reflexos do uso de deepfakes como prova em processos jurídicos e como aprimorar o sistema brasileiro nesse quesito. Assim, foi possível não só obter uma análise dos limites e desafios para o combate dos deepfakes, destacando os reflexos no âmbito penal inerente ao uso dessa tecnologia, mas como também verificar a ausência de preparo legal e tecnológico do sistema jurídico brasileiro para lidar com a democratização do acesso aos *deepfakes* e seus usos no processo penal.

Palavras-chave: I.A; *deepfakes*; provas; processo penal; tecnologia.

INTRODUÇÃO

A globalização veio trazendo diversas mudanças na sociedade, e uma delas foi o crescimento exponencial do uso da internet. A rápida evolução dessa tecnologia tem gerado profundas transformações sociais, incluindo o âmbito jurídico. Uma das preocupações que têm surgido é a disseminação dos chamados "*deepfakes*", que são vídeos, áudios e imagens criadas por meio da Inteligência Artificial - IA e que parecem extremamente reais, mas não passam de materiais fabricados de forma fraudulenta, gerando uma confusão e uma insegurança ao tentar diferir o que foi manipulado e a realidade. Essa questão traz uma problemática jurídica complexa, especialmente no que diz respeito à sua utilização na produção de provas em processos judiciais e às repercussões penais envolvidas.

No contexto atual, os deepfakes representam um desafio para a busca da verdade não só em um processo judicial mas como também na sociedade como um todo. A capacidade de manipular imagens e vídeos com precisão, de forma acessível e em larga escala, gera um grande desafio para o sistema jurídico. A habilidade de criar evidências audiovisuais fictícias levanta questões sobre a autenticidade das provas apresentadas em tribunal, e questiona a confiabilidade das próprias tecnologias de gravação, que historicamente eram tidas como incontestáveis. Logo, com o avanço das técnicas de manipulação digital, acaba ficando cada vez mais difícil distinguir o que é real e o que não é. E esse fato aborda questões importantes e necessárias sobre a confiabilidade das provas produzidas com o auxílio da Inteligência Artificial e como os tribunais e os profissionais do Direito devem lidar com esses desafios.

Além disso, à medida que a produção de deepfakes se torna mais acessível e sofisticada, surge a necessidade de abordar questões relacionadas à responsabilidade penal. Quem deve ser responsabilizado quando uma IA é usada para criar conteúdo enganoso com intenções fraudulentas, difamatórias ou criminosas? Como a legislação atual lida com essas situações e como ela deve evoluir para enfrentar esse desafio? O atual cenário tecnológico e jurídico possui ferramentas o suficiente para garantir que não haja equívocos processuais baseados em provas manipuladas por deepfakes?

Diante desses efeitos, este trabalho propôs uma análise aprofundada das implicações jurídicas dos deepfakes, considerando as nuances do uso da Inteligência Artificial na produção de provas legais e nas responsabilidades penais envolvidas. Serão examinados casos emblemáticos, legislações e debates em andamento sobre esse tema.

Portanto, o objetivo do presente artigo é esclarecer a extensão em que o fenômeno dos deepfakes consegue influenciar negativamente na produção de provas no sistema penal

brasileiro e, a partir disso, identificar lacunas legais e propor possíveis soluções para mitigar os efeitos negativos dos deepfakes no sistema jurídico.

1 PROVAS LÍCITAS E ILÍCITAS NO PROCESSO PENAL

1.1 Definição e classificação de provas no processo penal

Um dos princípios mais difundidos no âmbito do Processo Penal é o Princípio da Busca pela Verdade, princípio este que ainda está, de certa forma, atrelado ao que é amplamente denominado de "verdade real": a verdade real é aquela que estabelece que o julgador deve sempre buscar estar mais próximo das realidades ocorridas no fato, e determina também que o fato investigado deve corresponder ao que está fora dele, em toda sua plenitude, através de provas.

De acordo com o professor Rogério Lauria Tucci[1], o conceito de verdade real poderia ser descrito como "a reconstrução atingível de fato relevante e metaprocessual, inquisitivamente perquirida para deslinde da causa penal". Além disso, Tourinho Filho[2] afirma que "no Processo Penal o juiz tem o dever de investigar a verdade real, procurar saber como os fatos se passaram na realidade, que realmente praticou a infração e em que condições a perpetrou, para dar base certa à justiça". Dessa forma, entende-se que a verdade utilizada no Processo Penal deve refletir ao máximo a realidade dos fatos contemplados, objetivando então transmitir, de forma plena e fiel, essa realidade nos autos do processo. Por isso, no âmbito do processo penal, é comum ouvir que é impossível chegar a uma verdade absoluta dos fatos.

Nesse sentido, o doutrinador Aury Lopes Júnior[3] conceitua as provas como os meios pelos quais se torna possível a reconstrução aproximativa de um determinado fato histórico (crime), tendo como objetivo criar condições para que o juiz exerça sua atividade recognitiva, a partir da qual se produzirá o convencimento externado na sentença.

O Código de Processo Penal traz os tipos de provas aceitas pelo sistema jurídico, sendo elas: prova pericial, prova documental, prova testemunhal. Nesse sentido, Nucci[4] traz a seguinte classificação: prova pericial é aquela formada a partir de uma perícia por especialistas, como o

---

[1] TUCCI, R.L. **Princípios e regras orientadoras do Novo Processo Penal Brasileiro**. São Paulo: Forense, 1986. p. 145.
[2] FILHO, T. **Processo Penal**. V. 1. São Paulo, Ed. Saraiva, 2000, p.41.
[3] JR., Aury L. **Direito processual penal**. São Paulo: Editora Saraiva, 2023. Disponível em: https://integrada.minhabiblioteca.com.br/#/books/9786553626355/. Acesso em: 12 set. 2023.
[4] NUCCI, Guilherme de S. **Manual de Processo Penal**. São Paulo: Grupo GEN, 2022.. Disponível em: https://integrada.minhabiblioteca.com.br/#/books/9786559643691/. Acesso em: 12 set. 2023.

exame de corpo de delito por exemplo, onde há a verificação da prova da existência do crime, feita por peritos, diretamente, ou por intermédio de outras evidências, quando os vestígios, ainda que materiais, desapareceram. A prova documental é aquela que concentra elementos aptos a provar um fato ou acontecimento, baseada por exemplo em escritos, fotos, fitas de vídeo e som, desenhos, esquemas, gravuras, disquetes, CDs, entre outros. Por fim, a prova testemunhal é aquela feita de depoimentos prestados por testemunhas que viram e presenciaram o fato delituoso, ou até mesmo possuem uma informação importante sobre o caso.

1.2 Provas lícitas, ilícitas e a Teoria dos Frutos da Árvore Envenenada

Nucci[5] afirma que os meios de prova "são todos os recursos, diretos ou indiretos, utilizados para alcançar a verdade dos fatos no processo". A partir disso, podemos classificar as provas como lícitas ou ilícitas: as provas lícitas são todas aquelas que foram obtidas através de meios admitidos pelo ordenamento jurídico, enquanto as provas ilícitas são aquelas que foram obtidas por meios proibidos em lei ou até mesmo por meios imorais, antiéticos, atentatórios à dignidade e aos bons costumes, ou contrários aos princípios gerais de direito. Assim, neste caso, as provas ilícitas devem ser desentranhadas do processo, conforme dispõe o art. 157 do CPP.

Ainda no que concerne sobre as provas ilíticas, tem-se a Teoria dos Frutos da Árvore Envenenada: essa teoria deriva do Princípio da contaminação, e teve sua origem no caso Silverthorne Lumber & Co. v. United States, em 1920. Em suma, essa teoria defende que todas as provas decorrentes de prova ilícita são contaminadas por este vício, ou seja, acabam sendo consideradas ilícitas também.

Embora a teoria dos frutos da árvore envenenada seja comumente aplicada pelos tribunais brasileiros, incluindo o Superior Tribunal de Justiça (STJ), existem duas outras teorias que estabelecem limites para sua aplicação. Uma delas é a teoria da descoberta inevitável, que permite o uso de uma prova ilícita por derivação se for possível demonstrar que essa prova teria sido de qualquer forma descoberta por meios legais no curso normal da investigação. A outra é a teoria da fonte independente, que argumenta que uma prova derivada de uma fonte ilícita não deve ser descartada se também tiver uma origem legítima que não esteja relacionada à prova ilícita inicial.

---

[5] NUCCI, Guilherme de S. **Manual de Processo Penal**. São Paulo: Grupo GEN, 2022.. Disponível em: https://integrada.minhabiblioteca.com.br/#/books/9786559643691/. Acesso em: 12 set. 2023.

1.3 Regras de apreciação de provas e sua evolução

Os sistemas de valoração de provas são basicamente três, sendo eles o sistema de livre convicção, de prova legal e persuação racional: o sistema de livre convicção possui uma valoração livre, em que o julgador decidirá conforme sua íntima convicção baseada nas provas dos autos, provas que não estejam nos autos ou até mesmo em desacordo com as provas dos autos, não sendo necessário motivar sua decisão, como ocorre no Tribunal do Júri. O sistema da prova legal, por sua vez, entende que o valor probatório dos meios de prova é pré-determinado pelo legislador, e cabe ao juiz apenas avaliar o conjunto das provas e atribuir-lhes o valor conforme estabelecido na lei, sem margem para sua própria avaliação subjetiva do valor das provas. Por fim, o sistema de persuação racional é considerado um método misto, onde o juiz possui uma grande autonomia e liberdade na valoração das provas, desde que fundamente sua decisão. Esse é o sistema adotado majoritariamente pelo processo penal brasileiro, encontrando, inclusive, fundamento na Constituição Federal (art. 93, IX).

Nos últimos anos, o crescente avanço da tecnologia vêm alterando o cenário de avaliação e aceitação de provas. A digitalização de documentos e evidências digitais, por exemplo, influenciaram diretamente as regras de apreciação de provas e os desafios associados. A 5ª Turma do Superior Tribunal de Justiça, por exemplo, reconheceu a validade de reproduções de mensagens do WhatsApp como prova em processo penal desde que sejam observadas as regras de admissibilidade e autenticidade. Nesse sentido, é possível observar que nos últimos anos houve um crescimento na utilização de provas digitais, principalmente provas em fotos, vídeos e áudios, além do uso de inteligência artificial na produção destas provas. Portanto, a partir da evolução e adequação dos novos meios de prova aceitos pela lei, é de suma importância uma triagem e análise minuciosa da autenticidade de cada prova apresentada para compor o convencimento do magistrado e do julgador, para que seja possível chegar perto da verdade dos fatos e, logo, aplicar a lei de forma correta e efetiva.

2 DIREITO PENAL E TECNOLOGIA

2.1 Cibercrimes e delitos digitais

O direito penal e a tecnologia nem sempre andam de lados opostos. Um bom exemplo de caso em que esses dois elementos foram utilizados com excelência é o narrado com maestria

pelo juiz Alexandre Morais da Rosa[6]: conta-se de uma situação em que um homem foi acusado de roubo através do reconhecimento pela vítima, que fora rendida a mão armada. O advogado deste homem, para provar a inocência de seu cliente, fez com que este autorizasse a extração da "linha do tempo" de seu celular, já que haviam registros feitos com o e-mail do Google pelo próprio aparelho. Através disso, foi possível obter a localização do aparelho no momento do crime. Contudo, uma vez que essa linha do tempo pode ser facilmente editada, o delegado do caso verificou a localização no horário do roubo e determinou diligências de constatação nas câmeras de segurança dos condomínios próximos à avenida que, ao tempo do fato, o smartphone estava. Assim, foi possível a aquisição de imagens do veículo e do condutor que, no caso, era o homem injustamente acusado, excluindo a responsabilidade penal do mesmo. Assim, ao invés de focar na preparação da audiência de custódia ou na impugnação do dito reconhecimento ilegal, o advogado utilizou-se da tecnologia disponível para, de forma rápida e simplificada, provar a inocência de seu cliente. Contudo, essa dupla do direito penal e tecnologia nem sempre trouxeram boas mudanças na sociedade, principalmente nos últimos anos.

Ao longo dos últimos anos verificou-se que o aumento das atividades ilegais através de novas tecnologias de informação e comunicação está cada vez mais causando diversos prejuízos para a sociedade, incluindo atividades ilícitas relacionadas a fraudes financeiras, promoção de crimes, invasões de privacidade, pornografia infantil, entre outros. Geralmente os cibercrimes englobam atividades ilegais que ocorrem exclusivamente no espaço cibernético, como hacking, phishing, ataques de negação de serviço, etc. Por outro lado, os delitos digitais normalmente incluem atividades ilegais que envolvem o uso de dispositivos digitais, como a falsificação de documentos digitais, crimes financeiros online, ciberbullying e assédio via internet. Assim, ambos englobam atividades ilegais que exploram as vulnerabilidades e a interconexão da era digital. Infelizmente, esses crimes estão cada vez mais comuns e, embora uma parcela da sociedade já esteja ciente de como se proteger, a grande maioria desconhece muitos desses crimes e infelizmente continuam levando grandes prejuízos dos criminosos.

Com isso, o sistema jurídico enfrenta desafios consideráveis no combate aos cibercrimes e delitos digitais, vez que a maioria das leis penais foi concebida em uma era pré-digital e muitas vezes não está equipada para lidar com a complexidade desses crimes. A atribuição de jurisdição e a cooperação internacional também são desafios, uma vez que muitos desses crimes

[6] ROSA, Alexandre Morais da. **Quando o defensor e a tecnologia viram o jogo no flagrante.** Revista Consultor Jurídico, 2022. Disponível em: https://www.conjur.com.br/2022-nov-04/limite-penal-quando-defensor-tecnologia-viram-jogo-flagrante. Acesso em 13 out 2023.

transpassam fronteiras. Não obstante, a identificação e a autenticidade das provas digitais podem ser difíceis de estabelecer de forma irrefutável.

Os impactos desses tipos de delitos são vários, e vão desde prejuízos financeiros e danos à reputação a impactos na privacidade e consequências psicológicas e emocionais. Nesse sentido, o Estado e empresas privadas se viram diante de desafios para combater e minimizar os danos causados por delitos digitais, como por exemplo, o uso de antivírus e firewalls, constantes atualizações de software, políticas de segurança corporativa, criptografia de dados e, principalmente, conscientização e educação digital.

2.2 Legislação e regulamentação

A tecnologia vem modificando o Direito de modo significativo não apenas no Brasil, mas no mundo inteiro. A legislação foi obrigada a mudar e se reinventar a partir dos desdobramentos da tecnologia principalmente no âmbito penal. Com esse avanço, a legislação brasileira e os aplicadores da lei entraram em um debate acerca da aplicação ou não de normas penais existentes às novas condutas violadoras de direitos praticadas através das novas tecnologias. Nestes casos, observa-se que a tendência tem sido aplicar as leis penais existentes, mesmo que isso signifique estender a interpretação analógica para incluir as novas atividades ilegais facilitadas pelas novas tecnologias, enquanto as mudanças legislativas propostas avançam lentamente no atual ambiente legislativo nacional. Algumas leis que versam sobre essa relação do direito penal e da tecnologia no Brasil são : a) Lei Federal n.º 12.735/2012[7] (Lei Azeredo) para hipóteses de *deepfake* pornográfico; b) Lei Federal n.º 12.737/2012[8] (Lei Carolina Dieckmann); c) Código Eleitoral (Lei nº 4.737/1965[9]) para as situações envolvendo conteúdo político durante as eleições do Poder Executivo e d) Código Penal[10] para *deepfakes* caluniosos, difamatórios e injuriosos. Ainda, esses crimes podem ser amparados de maneira

[7] BRASIL. Lei Federal n.º 12.735, de 30 de novembro de 2012. Dispõe sobre a "responsabilização dos provedores de conexão e de aplicações de internet por danos decorrentes de conteúdo gerado por terceiros" e altera o art. 7º da Lei nº 8.137, de 27 de dezembro de 1990 (Código de Defesa do Consumidor), e o art. 109 da Lei nº 9.279, de 14 de maio de 1996 (Lei da Propriedade Industrial), para os fins que especifica. **Diário Oficial da União**, Brasília, DF, 3 dez. 2012. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12735.htm

[8] BRASIL. Lei Federal n.º 12.737, de 30 de novembro de 2012. Define os crimes cibernéticos e dá outras providências. **Diário Oficial da União**, Brasília, DF, 3 dez. 2012. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12737.htm

[9] BRASIL. Lei nº 4.737, de 15 de julho de 1965. Institui o Código Eleitoral. **Diário Oficial da União**, Brasília, DF, 16 jul. 1965. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l4737compilado.htm

[10] BRASIL. Decreto-Lei nº 2.848, de 7 de dezembro de 1940. Institui o Código Penal. **Diário Oficial da União**, Rio de Janeiro, RJ, 31 dez. 1940. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto-lei/del2848compilado.htm

geral na Lei de Crimes Cibernéticos (Lei nº 12.737/2012), no Marco Civil da Internet (Lei nº 12.965/2014 — "MCI"[11]) e na Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018[12]), que abordam a preservação de direitos civis e apenamento de infratores tanto no âmbito cível quanto criminal.

No entanto, essa relação entre o Direito Penal e Tecnologia exige que as leis estejam em constante mudança e desenvolvimento e, nesse sentido, verifica-se, por exemplo, a ausência de uma lei concreta que verse especificamente sobre o uso e limites da Inteligência Artificial. Felizmente, esse tema já está avançando na área jurídica: o deputado federal Eduardo Bismarck (PDT-CE) teve seu projeto de lei - PL 21/2020 aprovado no plenário, e visa estabelecer "*fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil*". Mesmo sendo um início, ainda não há a previsão de um conjunto de normas e regulamentos concretos que visem regular a utilização da inteligência artificial pela sociedade e, consequentemente, mitigar os riscos que ela pode gerar principalmente no âmbito penal.

3 FENÔMENO DO DEEPFAKE, SEUS REFLEXOS EM MATÉRIA PROBATÓRIA E COMO APRIMORAR O SISTEMA BRASILEIRO

3.1 Definição de deepfakes

O aumento dos usuários na internet tem uma consequência que está sendo objeto de discussão em vários países do mundo: a coleta e o armazenamento de dados. Essa transformação digital trouxe novas possibilidades do uso "inteligente" das grandes massas de dados, que é comumente utilizada por empresas e pelo Estado de diversas formas. A partir do controle e processamento desses dados, é possível criar diversos recursos, como por exemplo a Inteligência Artificial: a IA consiste em mecanismos computacionais que são baseados no comportamento humano para a realização de tarefas ou até mesmo para a resolução de problemas. Em suma, é como se o computador "pensasse" e executasse tarefas exatamente como um ser humano, baseando-se no processamento e interpretação dos dados em que ele tem disponível.

---

[11] BRASIL. Lei Federal n.º 12.965, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. **Diário Oficial da União**, Brasília, DF, 24 abr. 2014. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm

[12] BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet), e a Lei nº 9.472, de 16 de julho de 1997 (Lei Geral de Telecomunicações). **Diário Oficial da União**, Brasília, DF, 15 ago. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

Assim, um tema que vem ganhando bastante notoriedade nos dias atuais é a utilização dos deepfakes, uma ferramenta que utiliza recursos da Inteligência Artificial para a produção de conteúdo audiovisual das mais variadas maneiras possíveis. É através dela, por exemplo, que se torna possível a substituição realista de rostos em vídeos, a imitação de vozes que se assemelham muito à voz do locutor original e consequentemente uma distorção de seu conteúdo inicial, e etc. O uso dessa tecnologia vem sendo utilizada de diversas formas, desde a criação de vídeos humorísticos e conteúdos educacionais até a manipulação de vídeos com celebridades, políticos e etc. Por ser um instrumento de altíssima e sofisticada tecnologia, as *deepfakes* apresentam uma capacidade inestimável de "distorcer a verdade e por a risco a reputação dos envolvidos, criando um profundo prejuízo aos atingidos quando utilizado para fins escusos"[13]. Por isso, é uma técnica de manipulação de imagens e vídeos muito temida nessa nova era da tecnologia, uma vez que são utilizados recursos altamente capazes de confundir o receptor do conteúdo, de forma em que se torna difícil saber se o conteúdo de fato aconteceu no mundo real, ou não.

Ultimamente, as mídias vêm abordando cada vez mais sobre o assunto, com o intuito de levar o conhecimento dessa nova tecnologia para a sociedade, bem como para alertá-las sobre seus riscos. Contudo, os *deepfakes* vêm ganhando cada vez mais espaço nas mídias digitais e, consequentemente, aumentando o risco de espalhar desinformação para os usuários nas redes. Não só isso, esse recurso acabou gerando repercussões no âmbito jurídico do mundo todo, levantando diversos questionamentos nos quais muitos ainda não têm resposta consolidada em leis ou jurisprudências.

3.2 Técnicas utilizadas na produção de deepfakes e casos de uso

Como já explicado, as *deepfakes* são conteúdos de vídeos, áudios e imagens extremamente realistas e semelhantes ao original, porém, obviamente, falsos. Esse recurso vem sendo mais abordado nos últimos anos e, conforme Moraes[14]:

> Em questão de datas, o primeiro ocorreu no Outono de 2017 utilizado para gerar conteúdos adultos. Posteriormente, essa técnica foi melhorada por uma pequena

[13] MOLINA, A, C.; BERENGUEL, O, L. **Deepfake: A evolução das fake news**. Research, Society and Development, v. 11, n. 6, e56211629533, 2022. DOI: http://dx.doi.org/10.33448/rsd-v11i6.29533. Disponível em: https://rsdjournal.org/index.php/rsd/article/view/29533. Acesso em 10 jul 2023.

[14] MORAES, Cristiane Pantoja De. **"Deepfake" como ferramenta de manipulação e disseminação de "fakenews" em formato de vídeo nas redes sociais.** 2020. Biblios, ISSN 1562-4730 No 79 (2020). DOI 10.5195/biblios.2020.864. Disponível em: https://dialnet.unirioja.es/descarga/articulo/8041632.pdf. Acesso em 14 jul 2023.

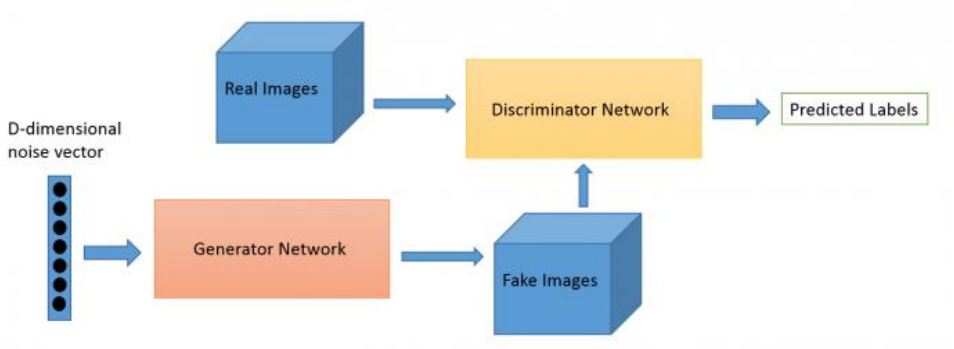
> comunidade para criar nomeadamente uma aplicação chamada “FakeApp”. O processo para gerar deepfake consiste em imagens que reúnem rostos alinhados de duas pessoas diferentes, nas quais há a reconstrução do rosto de uma em conjunto de dados de imagens faciais das outras e se auto codifica para então reconstruir rostos com as imagens faciais. Na prática, os resultados são impressionantes, o que explica a popularidade da técnica. O último passo é levar o vídeo ao alvo, extrair e alinhar a face do alvo a partir de cada quadro, utilizando software ou aplicativos “FaceApp”.

Antes de compreender as técnicas utilizadas na produção de deepfakes em si, é fundamental conhecer o berço de onde nasceu: *Big Data* e *Machine Learning*. A *Big Data* é o termo em que se trata dos grandes conjuntos de Dados que são processados, armazenados e analisados, para que seja possível atribuir significado a eles e, a partir disso, traçar estratégias e ações. É um instrumento de análise de dados muito utilizado por bancos, empresas, comércio e até mesmo pelo governo. Enquanto isso, o *Machine Learning* é o conjunto de técnicas, geralmente criadas pelos estudos da Engenharia, Estatística e Ciências da Computação, que busca capturar padrões comportamentais de uma base de dados e ensiná-los a uma máquina. É considerado uma vertente da Inteligência Artificial, e está diretamente ligado com a capacidade que uma máquina tem de tomar decisões a partir de um raciocínio similar ao do pensamento humano.

Para se criar um conteúdo baseado em um *deepfake*, um dos meios mais utilizados são baseados em GANs: Redes Generativas Adversárias (*Generative Adversarial Networks* - GANs), uma classe da Inteligência Artificial inventada por Ian Goodfellow, em 2014, e que vem sendo utilizada não só em criação de conteúdo visual, mas como também na área de estudos do Marketing, Finanças, Economia, etc. Em suma, esse instrumento é capaz de produzir novo conteúdo a partir do zero, é constituído por dois algoritmos principais: o discriminador e o gerador. O algoritmo discriminador tem o objetivo de “classificar e categorizar os dados de entrada”, enquanto os algoritmos geradores, por sua vez, preveem os recursos ou os destinos com as determinadas categorias rotuladas pelo algoritmo discriminador.

A partir disso, a rede neural conhecida como gerador é utilizada para criar novos dados, enquanto outra rede neural, o discriminador, é responsável por avaliar se esses novos dados são genuínos ou não, ou seja, se parecem pertencer ao conjunto de dados originais utilizados para treinar a rede. Esse processo pode ser ilustrado conforme a figura 1:

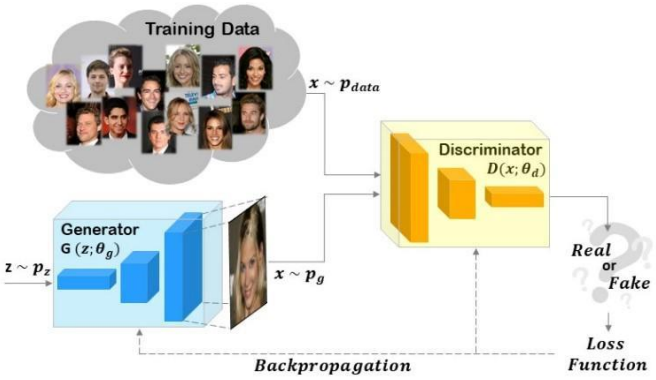
Figura 1 - Processo de criação dos deepfakes:



Fonte: Deep Learning Book, 2022.

Como o próprio nome sugere, a relação entre esses dois componentes é feita de forma "adversária", no sentido de que, enquanto o discriminador é treinado para diferenciar o real do falsificado, o gerador é treinado para enganar o discriminador com o conteúdo que está produzindo. Por isso, as GANs são conhecidas por serem dois sistemas derivados da Inteligência Artificial que aprendem ao competir uma com a outra. Assim, com o processo de treinamento de ambos que vão se desenvolvendo em conjunto, têm-se resultados cada vez mais perfeitos:

Figura 2 - Formação do conteúdo com dados disponíveis:



Fonte: Journal of Imaging, 2022.

Embora as GANs sejam eficientes na síntese de imagens, elas apresentam dificuldades em manter a consistência temporal e a coerência entre os quadros em vídeos, tornando-as inadequadas para criar deepfakes em vídeo. Contudo, muitos desses obstáculos já não são mais obstáculos, devido ao desenvolvimento e aprimoramento constante dessa ferramenta. Além disso, na criação de audiofakes, as GANs não são tão utilizadas. Na maioria dos casos, a produção de deepfakes envolve uma variedade de algoritmos de inteligência artificial que trabalham juntos de forma conjunta e interdisciplinar.

O site "Explicando IA", projeto desenvolvido pelo Oxford Internet Institute em parceria com a Google, trouxe um exemplo interessante: suponhamos que a tarefa designada seja criar

uma imagem baseada no estilo artístico de Pablo Picasso. Nesse caso, foram aplicados todos os dados disponíveis sobre o estilo do artista, para que um algoritmo produza milhares de novas imagens no estilo de Picasso enquanto o outro analisa criteriosamente cada tentativa para diferenciar as que obtiveram sucesso e as que falharam. O resultado encontrado foi esse:

Figura 3 - Imagem criada por IA:

Fonte: Explicando IA, 2022.

A partir da compreensão dos GANs, é possível entender como são criados os conteúdos dos deepfakes. Como visto anteriormente, as GANs têm a capacidade de executar a grande maioria das tarefas que lhe são atribuídas, e uma das demandas que vêm sendo demandadas à elas são casos em que há a manipulação em rostos humanos, tanto para criar um rosto "do zero" (*Entire Face Synthesis*) ou fazer mudanças em um rosto já existente *(Attribute manipulation*) quanto para colocar o rosto de uma pessoa no corpo de outra em um vídeo extremamente realista. Por isso, as deepfakes podem ser classificadas de quatro formas: face replacement, face reenactment, face generation e audio synthesis.

Ao contrário do que muitos pensam, essa tecnologia não é extremamente difícil de ser utilizada por pessoas leigas no assunto. Ao mesmo tempo em que ela está avançada, essa tecnologia está sendo cada vez mais simplificada para que seja mais fácil seu uso, e um dos exemplos para isso é o aplicativo de celular denominado de FaceApp. Disponível na App Store, o FaceApp é um aplicativo que vem ganhando cada vez mais notoriedade principalmente devido à facilidade que o usuário tem em utilizar os instrumentos baseados na inteligência artificial só apertando alguns botões. Ele possibilita que, a partir do upload da foto de alguém, seja possível alterar seu rosto, cabelo, sexo e até idade de forma realista usando apenas o telefone celular:

Figura 4 - Criação de DeepFake simples pelo aplicativo FaceApp:

Fonte: Expresso PB, 2022.

Em relação aos chamados “face replacements”, que consiste na troca de rostos em um corpo, por exemplo, o ciclo geral de criação se dá da seguinte maneira, de acordo com o MIT Technology Review:

> 1 – A rede neural detecta e recorta a face da imagem que será acoplada no conteúdo falso;
> 2 – Técnicas específicas são usadas para separar as expressões do rosto para em seguida modificá-las;
> 3 – A representação intermediária do rosto/corpo é convertida na expressão desejada para a geração, isso antes de renderizar a imagem;
> 4 – Gera-se uma nova face com base em algum ponto focal (o rosto da imagem que serviu como base);
> 5 – A composição do conteúdo original é importante (cabelo, cena, etc) combinado com uma renderização 3D, imagem distorcida do conteúdo gerado, etc., combinado com técnicas que refinam o realismo;
> 6 – Combina-se a face gerada e a cena de destino, para assim compartilhar o conteúdo falso;

Já para áudios, normalmente, a técnica utilizada é fragmentar o áudio em partes menores e, para cada uma delas, calcular os Coeficientes Mel-Cepstral (MCC), que conseguem captar as frequências da voz mais presentes. Esses dados são combinados posteriormente com as imagens geradas pelo vídeo.

3.3 Exemplos de *deepfakes*

Para ilustrar melhor a precisão da tecnologia apresentada, segue alguns exemplos de imagens e recortes de vídeos criados com esse instrumento:

Figura 5 - DeepfakeTIMIT é um banco de dados de vídeos em que rostos são trocados usando a abordagem baseada em GAN (Redes Adversárias Generativas) de código aberto.

Fonte: Idiap Research Institute, 2022.

Figura 6 - utilização do deepfake para alterar a expressão facial através do método Face2Face.

Fonte: IEEE Conference on Computer Vision and Pattern Recognition, 2016.

Figura 7 - capturas de tela do vídeo "*Você não vai acreditar no que o Obama disse!*", em que foram substituídas as palavras do ex-presidente dos EUA Barack Obama pelas do ator e cineasta Jordan Peele.

Fonte: BuzzFeed News, 2018.

Esse vídeo foi criado pela equipe do BuzzFeed News com o objetivo de mostrar como é fácil fazer essa manipulação e, por outro lado, como é difícil diferenciar se ela de fato é real ou não, tendo em vista o alto realismo das imagens e da alteração da voz. Segue a transcrição do vídeo deepfake realizado[15]:

[15] SILVERMAN, C. Como identificar um "deepfake" como este vídeo do Barack Obama. **BuzzFeed News,** 2018. Disponível em: https://www.buzzfeed.com/br/craigsilverman/como-identificar-deepfake-video-obama-peele. Acesso em: 28 jul 2023.

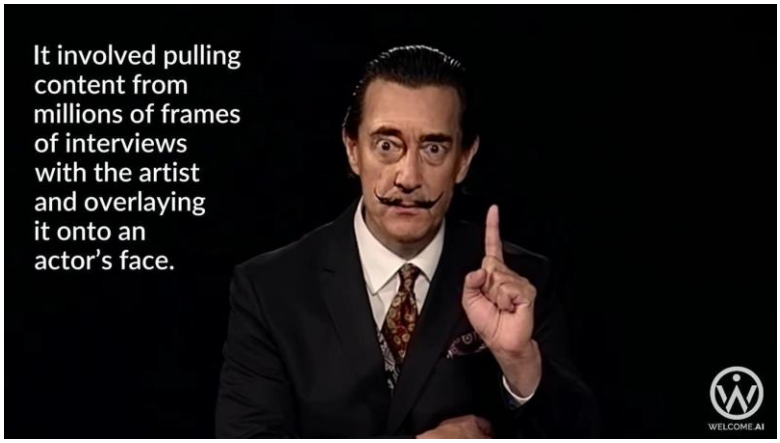
> Estamos entrando em uma Era na qual nossos inimigos podem fazer com que qualquer um pareça estar dizendo qualquer coisa a qualquer momento, mesmo que eles nunca tenham dito isso. Então, por exemplo, poderiam me fazer dizendo coisas como… Não sei… "Killmonger estava certo", ou, "Ben Carson está no lugar profundo do 'Corra!', ou, que tal, simplesmente: "O presidente Donald Trump é um total e completo merda". Agora, vejam, eu nunca diria essas coisas. Pelo menos não em um discurso público, mas outra pessoa o faria. Alguém como Jordan Peele. Este é um momento perigoso. Ao avançar precisamos ficar atentos com aquilo que acreditamos na internet. É uma época em que precisamos contar com fontes confiáveis de notícias. Pode soar básico, mas a forma com a qual avançamos na Era da Informação fará a diferença entre nós sobrevivermos, ou nos tornarmos um tipo de distopia. Obrigado, e fiquem atentos.

Como foi possível observar, a técnica usada para criação dos deepfakes está sendo cada vez mais desenvolvida para que chegue o mais próximo possível da realidade. A seguir, ilustramos alguns casos de uso de deepfakes que vêm confundindo os internautas que os vêem.

Figura 8 - "Tom Cruise no Tiktok": o especialista de efeitos visuais Chris Ume, que está por trás do perfil "@deeptomcruise" na rede social TikTok se dedica à criação de vídeos extremamente realistas em que, através do deepfake, simula ser o próprio ator.

Fonte: Youtube, 2021.

Figura 9 - "Revivendo Dali".



Fonte: Youtube, 2019.

Em 2019, para celebrar o aniversário de 115 anos do artista espanhol Salvador Dalí, o museu dedicado a ele na cidade de São Petersburgo, Flórida (EUA), criou um deepfake do artista para interagir com os visitantes. A tecnologia de inteligência artificial usada para criar o deepfake foi alimentada com cerca de 6 mil frames de filmes com o próprio Dalí e mais de 190 mil combinações de diálogos.

De fato, o deepfake é um recurso bastante interessante para a criação de conteúdo na área do entretenimento. Tem um potencial muito grande, principalmente em relação à indústria cinematográfica, que pode obter grandes benefícios com o aprimoramento dessa tecnologia.

Contudo, nem sempre ela foi usada para entreter os seus telespectadores. Em alguns casos, ela foi utilizada para desinformar, confundir, enganar e até mesmo para praticar crimes.

Figura 10 - Captura de tela de vídeo deepfake mostrando o presidente ucraniano Volodymyr Zelensky:

Fonte: Youtube, 2022.

Nesse vídeo, é mostrado um deepfake do presidente ucrniano anunciando uma redução nas tropas da Rússia, e foi compartilhado em redes sociais e postado em um site de notícias local por hackers. No entanto, o vídeo foi rapidamente retirado pelas principais plataformas após ser identificado como falso. Assim, tem-se uma noção da periculosidade desse instrumento da IA ao ser usado para fins políticos.

Entrando na área que concerne aos crimes cibernéticos, uma prática que está sendo amplamente usada através do *deepfake* é a manipulação de vídeos pornográficos para colocar o rosto de determinadas pessoas, principalmente mulheres, no corpo do ator ou atriz de conteúdo

adulto. Um caso recente foi o ocorrido com a atriz Scarlett Johansson e, de acordo com o site "Filmow"[16]:

> Scarlett Johansson está plenamente consciente de que existem vídeos pornôs com seu rosto, alguns dos quais acumulam milhões de visualizações, mas ela recentemente disse ao The Washington Post que não há nada que possa realmente fazer sobre isso.
> "Claramente isso não me afeta tanto, porque as pessoas assumem que não sou eu em um filme pornô, por pior que seja", disse Johansson. "Eu acho que é uma luta inútil, legalmente, principalmente porque a internet é um vasto buraco de escuridão que se consome. Há muito mais coisas perturbadoras na teia escura do que isso, infelizmente. Eu acho que cabe a um indivíduo lutar por seu próprio direito à sua imagem, reclamar danos, etc. "
> "Cada país tem seu próprio direito legal em relação ao direito à sua própria imagem, então, embora você possa derrubar sites nos EUA que usam seu rosto, as mesmas regras podem não se aplicar na Alemanha", continuou a atriz. "Mesmo que você tenha imagens com direitos autorais e imagens que pertençam a você, as mesmas leis de direitos autorais não se aplicam no exterior. Eu tenho infelizmente estado nesta estrada muitas e muitas vezes. O fato é que tentar se proteger da internet e sua depravação é basicamente uma causa perdida, em sua maior parte ".

Um outro caso ocorrido nos Estados Unidos foi o de uma mãe que usou desse recurso para "favorecer" a filha em relação a garotas de um grupo de líderes de torcida. De acordo com o site de notícias estadunidense "The Philadelphia Inquirer[17]", uma mãe de uma das líderes de torcida teria compartilhado vídeos manipulados comprometedores, onde as supostas garotas apareciam nuas, fumando ou bebendo, comportamentos que não são bem vistos para líderes de torcida.

> O caso aconteceu em Chalfont, no estado da Pensilvânia, e os primeiros relatos datam do mês de junho de 2020. Segundo a polícia, uma das vítimas entrou em contato sobre uma imagem supostamente sua que recebeu de um contato, alegando que embora tivesse o seu rosto, não era verdadeira. A denúncia fez com que outras garotas também apresentassem queixas similares.
> A investigação procurou rastrear a origem das fotos e chegou até um site de telemarketing, de onde elas teriam sido disparadas para os números registrados, e ao cruzarem as informações, chegaram a um endereço IP comum como a fonte das imagens, que apontava para a residência de Raffaela Spone, moradora do mesmo distrito e que "coincidentemente" era mãe de uma das garotas do time... que não havia sido um alvo das fotos. […]
> O relatório alerta que o deepfake pode ser usado não apenas como uma ferramenta de bullying, mas também como uma arma contra indivíduos em posições de poder, incluindo líderes políticos, que podem ser desacreditados, ou pior, por conta de um vídeo falso em que aparecem dizendo e fazendo coisas controversas, ou mesmo cometendo crimes. Na época, os pesquisadores já haviam mencionado o óbvio poder do deepfake de destruir reputações, como as jovens que foram alvos de Spone.

[16] SCARLETT JONHANSSON FALA SOBRE DEEPFAKE EM VÍDEOS PORNOGRÁFICOS. **Filmow.** 2019. Disponível em: https://filmow.com/noticias/29534/scarlett-johansson-fala-sobre-deepfake-em-videos-pornograficos/. Acesso em: 2 ago 2023.

[17] VELLA, V,. A Bucks County woman created 'deepfake' videos to harass rivals on her daughter's cheerleading squad, DA says. **The Philadelphia Inquirer**, 2021. Disponível em: https://www.inquirer.com/news/bucks-county-raffaela-spone-cyberbullying-deepfake-20210312.html. Acesso em 9 ago 2023.

Além disso, é a partir desse recurso que muitos criminosos fazem o chamado "*revenge porn"*, que consiste no compartilhamento de fotos e vídeos íntimos das vítimas com a intenção de humilhação ou vingança. Com os recursos da inteligência artificial, há o favorecimento desse tipo de prática, uma vez que não é necessário ter, de fato, um conteúdo intimido produzido pela vítima. Há apenas a necessidade de manipulação de um determinado vídeo fazendo as alterações necessárias para que eles pareçam realistas. De acordo com o site do escritório de advocacia Favaretto Araújo Abreu[18]:

> Revenge porn ou pornografia de vingança é a expressão usada para denominar o ato de divulgar, na internet, fotos ou vídeos íntimos de terceiros, sem o consentimento dos mesmos.
> Casos como esse costumam acontecer, na maioria das vezes, quando um casal termina o relacionamento e uma das partes divulga as cenas íntimas na rede mundial de computadores, com o objetivo de vingar-se, ao submeter o ex-parceiro a humilhação pública.
> O vazamento de conteúdo íntimo traz diversas consequências à vítima. Já foram registrados vários casos de jovens que não aguentaram a exposição e cometeram suicídio. Quando não leva a atitudes extremas, o revenge porn deixa marcada a reputação de quem foi exposto. Isto quando não leva a problemas ainda mais sérios, que ultrapassam a esfera da moral, chegando a casos de agressões físicas e assédio sexual.

Ainda em relação aos crimes cibernéticos, ressalta-se a criação dos chamados "*deep nudes": uma* ferramenta do setor dos *deepfakes* que usa os algoritmos da inteligência artificial para criar nudes de qualquer mulher a partir de uma simples foto, mesmo que ela esteja completamente vestida. Há diversos sites abertos que proporcionam esse tipo de montagem, facilitando cada vez mais o mau uso desse recurso que, inclusive, não funciona com homens.

Figura 10 - Exemplo de imagem manipulada pela ferramenta Deep Nude:

Fonte: Aberto até de madrugada, 2019.

---

[18] O QUE É REVENGE PORN OU PORNOGRAFIA DE VINGANÇA?. **Favaretto Araújo Abreu Advogados,** 2022. Disponível em: https://favarettoadv.com.br/2023/08/10/o-que-e-revenge-porn-ou-pornografia-de-vinganca/. Acesso em 31 jul 2023.

De acordo com o site Canaltech[19]:

> O DeepNude funcionava por meio do download de uma ferramenta, onde algoritmos de inteligência artificial eram aplicados para atuar da seguinte forma: por meio do upload de uma foto de mulher — qualquer foto, vestida ou não —, o software tentaria criar uma imagem aproximada do corpo daquela pessoa nua. A qualidade dos resultados ficava aquém do esperado, com detalhes como borrões de imagem e granulações estranhas permeando o "falso nude". Mais além, o app adicionava marcas d'água na foto resultante para garantir que, caso a foto fosse espalhada por aí, ficasse claro que se tratava de uma foto fake.

A partir dos casos acima, é possível perceber que, mesmo tendo fins positivos em sua ideia inicial, os *deepfakes* podem ser usados também para fins diversos como desinformação, produção de conteúdo enganoso e principalmente para crimes. A polícia europeia Europol prevê que esse recurso da inteligência artificial será intensificado pelo crime organizado e que, por isso, é necessário que os órgãos da lei e seus servidores aprimorem suas habilidades e tecnologias para acompanhar o uso da nova técnica pelos criminosos.

De acordo com o Observatório do Laboratório de Inovação da Europol[20], as *deepfakes* estão sendo majoritariamente usadas em três áreas principais:

> 1) Desinformação: a Europol deu vários exemplos de como informações falsas podem ser divulgadas usando deepfakes, levando a consequências potencialmente devastadoras. Isso inclui na esfera geopolítica, como a criação de um falso alerta de emergência que avisa sobre um ataque iminente. Em fevereiro, antes do conflito Rússia-Ucrânia, os Estados Unidos acusaram o Kremlin de um complô de desinformação para servir de pretexto para uma invasão da Ucrânia.
> A tecnologia também pode ser usada para direcionar negócios, como criar um deepfake de vídeo ou áudio que faça parecer que o executivo de uma empresa está envolvido em um ato controverso ou ilegal. Em um caso bem divulgado, os criminosos fraudaram uma empresa de energia no valor de US$ 243 mil depois de se passarem pela voz do presidente-executivo.
> 2) Pornografia não consensual: O relatório citou um estudo da Sensity, que descobriu que 96% dos vídeos falsos envolviam pornografia não consensual. Isso normalmente envolve a sobreposição do rosto da vítima no corpo de um ator pornográfico, fazendo parecer que a vítima está participando do ato.
> 3) Fraude de documentos: embora os passaportes estejam se tornando cada vez mais difíceis de falsificar devido às modernas medidas de prevenção de fraudes, o relatório descobriu que "mídia sintética e imagens faciais manipuladas digitalmente apresentam uma nova abordagem para fraude de documentos". Por exemplo, essas tecnologias podem combinar ou transformar os rostos da pessoa a quem o passaporte

[19] ARBULU, R. App DeepNude, usado para "criar" nudes femininos, é descontinuado. **CanalTech**, 2019. Disponível em: https://canaltech.com.br/internet/app-deepnude-usado-para-criar-nudes-femininos-e-descontinuado-142863/.

[20] CRIME ORGANIZADO INTENSIFICARÁ USO DE DEEP FAKES, PREVÊ EUROPOL. **Ciso Advisor,** 2022. Disponível em: https://www.cisoadvisor.com.br/crime-organizado-intensificara-uso-de-deepfakes-preve-europol/. Acesso em 04 ago 2023.

> pertence e da pessoa que deseja obter um passaporte ilegalmente, aumentando as chances de a foto passar por verificações de identidade, incluindo as automatizadas.

Com isso, é notável que os *deepfakes* podem atrapalhar, e muito, o âmbito jurídico, principalmente no que concerne ao processo legal. Com o avanço dessa tecnologia, ficará cada vez mais fácil manipular ou criar provas para culpar ou retirar a culpa de alguém. Recentemente houve, por exemplo, um caso em que uma mãe manipulou uma gravação de áudio de seu marido com o objetivo de convencer o tribunal de que ele havia se comportado agressivamente em relação à sua filha. Ainda de acordo com o site CISO Advisor[21]:

> Para lidar eficazmente com esses tipos de ameaças, a Europol disse que os órgãos da lei devem desenvolver novas habilidades e tecnologias. Isso inclui detecção manual, que envolve a procura de inconsistências, e técnicas de detecção automatizadas, incluindo software de detecção de deepfake usando inteligência artificial que está sendo desenvolvido por organizações como o Facebook e a empresa de segurança McAfee.

Assim, nota-se que a democratização dessa tecnologia tem o poder de afetar gravemente a confiabilidade das provas apresentadas em um processo e, como consequência, a confiabilidade do sistema jurídico como um todo, como será demonstrado a seguir.

3.4 Análise da problemática relativa à utilização de deepfakes como prova em processos judiciais

Os *deepfakes* representam uma ameaça substancial no combate à desinformação. A capacidade de criar vídeos falsos convincentes, nos quais pessoas reais parecem estar dizendo ou fazendo coisas que nunca fizeram, pode ser explorada para manipular a opinião pública, uma vez que essa tecnologia não possui regulamentação própria ou limites nos dias atuais. Políticos, celebridades e figuras públicas podem ser alvo de campanhas difamatórias que prejudicam suas reputações e influenciam a percepção das pessoas sobre eles. Dessa forma, essa manipulação pode ter consequências graves para a democracia e o funcionamento saudável da sociedade.

O impacto dos *deepfakes* na desinformação e na confiança pública vai além dos danos individuais. A disseminação de conteúdo falso pode causar divisões sociais e políticas, como conseguimos ver nos dias atuais. *Deepfakes* podem ser usados, por exemplo, para fabricar

[21] CRIME ORGANIZADO INTENSIFICARÁ USO DE DEEP FAKES, PREVÊ EUROPOL. **Ciso Advisor,** 2022. Disponível em: https://www.cisoadvisor.com.br/crime-organizado-intensificara-uso-de-deepfakes-preve-europol/. Acesso em 04 ago 2023.

declarações controversas atribuídas a pessoas influentes, alimentando a polarização e o discurso de ódio de forma ilimitada. A desconfiança generalizada também pode levar ao surgimento de teorias conspiratórias e ao enfraquecimento do senso coletivo de realidade. À medida que a tecnologia avança, torna-se cada vez mais difícil distinguir entre vídeos reais e *deepfakes*, e isso leva a uma crise de confiança, onde as pessoas questionam a autenticidade de qualquer conteúdo audiovisual.

No que se refere aos crimes decorrentes do uso de deepfakes, a lista pode ser extensa, uma vez que as possibilidades são enormes. A tecnologia de hoje está avançada no nível em que é possível criar uma pessoa nas redes sociais a partir de fotos e vídeos manipulados a ponto de enganar perfeitamente seus usuários. Em uma situação como essa, a pessoa que criou essa "pessoa" estará cometendo um crime? Se um dos seguidores dessa pessoa desenvolve uma obsessão por essa pessoa criada e começasse a persegui-la e ameaçar sua "integridade física e psicológica", ele poderia ser enquadrado no crime de stalking?

Em outra situação, uma mulher acusa seu ex namorado de ameaça. Ela afirma que o homem a ameaçou na última vez que se viram e apresenta uma prova por um áudio em que ela "gravou escondido" no momento do crime quando, na realidade, ela apenas criou um áudio manipulado por "*deep voice*" para parecer que é seu ex namorado falando. Como o homem vai conseguir se defender? A justiça conseguiria analisar o material e verificar sua autenticidade? A mesma coisa pode acontecer com figuras políticas: um político pode ser investigado por corrupção se a única prova contra ele é um áudio de conversa gravada na qual ele admite a corrupção, onde na realidade o áudio é manipulado? Um indivíduo pode ser condenado por homicídio quando há um vídeo extremamente realista em que o rosto da pessoa foi trocado com o da pessoa que de fato matou?

Ainda, não será incomum ouvir casos em que, numa tentativa de incriminação, a vítima bebe demais, perde a consciência e no dia seguinte ela é acusada de um crime que não cometeu por conta de uma prova adulterada por *deepfake,* como por exemplo, um áudio gravado por alguém dele ameaçando sua esposa. Como a vítima poderá se defender ou apresentar algum argumento se nem ela se lembra do que fez?

Imaginemos também um caso onde há uma tentativa de incriminação de um indivíduo pelo crime de pedofilia, baseado num vídeo manipulado por deep fake e implantado nas redes da vítima. Ou até mesmo uma situação em que uma pessoa manipula um áudio ou vídeo e leva para delegacia com o intuito de obter uma medida protetiva contra o suposto agressor a fim de mantê-lo afastado de seu lar. Em casos assim, onde o procedimento deve ser mais ágil e célere

visando a segurança da vítima, a instituição estatal teria condições de filtrar e conferir prova por prova? Ainda, a instituição teria recursos para isto?

3.5 Análise dos limites e desafios legais, tecnológicos e sociais para o combate aos deepfakes e possíveis alternativas para prevenir e combater seu uso indevido

A partir do exposto, nota-se que a ascensão do fenômeno digital da Inteligência Artificial bota em cheque a confiabilidade de todo um sistema jurídico, não só do Brasil, mas do mundo todo. A tecnologia envolvida em todo processo está em constante evolução, dificultando as soluções de detecção dos deepfakes acompanharem essas mudanças e se manterem eficazes: à medida que os algoritmos de criação melhoram, fica cada vez mais difícil para os algoritmos de detecção identificarem as diferenças que separam o material falso do autêntico, aumentando então a probabilidade de espectadores comuns e até mesmo especialistas se equivocarem com o material espalhado. Dessa forma, é nítido o perigo que essa tecnologia possui e que pode acarretar inúmeras consequências principalmente no âmbito jurídico.

Uma reportagem feita pelo The Wall Street Journal[22] em 2019 relatou como uma empresa com sede no Reino Unido tomou um prejuízo de $243.000,00 devido a um golpe em que foi utilizado um tipo de deepfake, o chamado deep voice: o CEO da empresa de energia acreditou que estava ao telefone com seu superior, o executivo-chefe da empresa, quando seguiu as instruções para transferir imediatamente $243.000 para a conta bancária de um fornecedor húngaro. Contudo, a voz pertencia a um fraudador que utilizou a tecnologia de voz baseada em IA para falsificar a voz do executivo-chefe. Utilizando esse caso como exemplo, imaginemos: se o fraudador também tivesse manipulado mais materiais em que tal executivo-chefe falsamente admitisse a atitude criminosa, será que o fechamento deste caso seria diferente? Será que o judiciário teria recursos o suficiente para averiguar se tal confissão se tratava de uma manipulação ou não? Se fosse um caso julgado por tribunal do júri, por exemplo, a mera apresentação de uma prova aparentemente convincente pode ter um impacto devastador nos jurados e no resultado do julgamento.

Em sistemas jurídicos onde conversas por aplicativos de mensagens são aceitas como provas em processos judiciais, os deepfakes podem causar transtornos inimagináveis caso o Estado não tenha recursos o suficiente para averiguar a autenticidade do material. Se uma

[22] DAMIANI, J. A Voice Deepfake Was Used To Scam A CEO Out Of $243,000. **Forbes**, 2019. Disponível em https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/?sh=1d4d27b62241. Acesso em 06 ago 2023.

pessoa manipula um áudio de outra para parecer que esta lhe enviou um áudio por Whatsapp a ameaçando ou confessando um crime, por exemplo, seria viável prendê-la preventivamente?

Nesse sentido, verifica-se que há questionamentos a serem respondidos e, consequentemente, leis e procedimentos que trabalham esse tema: Como saber se uma prova apresentada em juízo pode ser realmente considerada verídica considerando a altíssima tecnologia disponível para a manipulação de material? O sistema judiciário de fato está preparado para os riscos que essa tecnologia pode acarretar? É possível a criação de normas que combatam os deepfakes sem censurar ou restringir a liberdade de expressão? A quem pertence os direitos autorais do deepfake? Haverão restrições de recursos das empresas e governos na implementação de medidas de combate aos deepfakes? Se o Estado interferir diretamente na produção de provas, teremos um autoritarismo digital?

O sistema legal, ao lidar com a variedade de novos comportamentos que ocorrem por meio dessas novas tecnologias, não pode ignorar essas questões. Muitos operadores do direito argumentam que as leis penais existentes estão desatualizadas para lidar com a eficácia da responsabilização dessas atividades ilegais, tendo, portanto, uma necessidade de criação de normas específicas que acompanhem especificamente tais atividades na medida que vão evoluindo e prejudicando inúmeras pessoas afetadas pela mesma.

Com a crescente democratização dessa tecnologia, o mundo se viu em uma situação de urgência não apenas para identificar o material manipulado do original, mas como também de criar soluções e procedimentos para agir diante de situações quando seu uso envolve má-fé. Por isso, vêm se falando muito sobre soluções para mitigar o uso indevido de *deepfakes*, tais como: marcação e autenticação de conteúdo através da blockchain para garantir a integridade das provas e mídias compatilhadas; educação e conscientização pública para ensinar a população a identificar sinais de manipulação; coleta de dados biométricos para autenticar a origem das informações, e etc.

No âmbito judicial, contudo, há uma demanda mais urgente em relação a identificação de deepfakes introduzidas como prova, visto que a mera apresentação de uma prova manipulada convinvente pode alterar todo o seu resultado, caso o sistema não consiga identificar prontamente que se trata de material adulterado. Algumas alternativas buscadas pelo sistema jurídico internacional para amenizar esse risco são: implementação de um sistema de autenticação e certificação de provas rigoroso para provas audiovisuais; investimento em tecnologias de detecção de *deepfakes;* inclusão de peritos em tecnologia e análise forense para o fornecimento de análises da autenticidade do material; estabelecer uma cadeia de custódia digital; introdução de normas específicas acerca do uso de *deepfakes* como prova e aplicando

uma pena mais rígida para quem se utilizar desse material com má-fé; fornecer educação jurídica para juízes, advogados, policiais e outros profissionais para que consigam identificar com mais precisão uma prova manipulada; criação de um mecanismo de auditoria independente que analisa a autenticidade das provas apresentadas e, por fim, um julgamento baseado em múltiplas fontes de prova além das mídias digitais para que seja reduzida a dependência de evidências audiovisuais.

Entrando em uma análise mais específica, verifica-se a necessidade de abordar a insuficiência da cadeia de custódia e seu impacto no âmbito digital. Devido à natureza volátil e delicada dos dados que servem como evidência, é possível que tanto os envolvidos no processo quanto os especialistas possam, intencional ou inadvertidamente, alterar, editar, manipular ou destruir evidências digitais. A "e-evidência" abrange tanto os formatos físicos quanto os lógicos e, desde a coleta até a eliminação, é essencial garantir a "identificação" dos dispositivos (tanto externos, como unidades de armazenamento, quanto internos, ou seja, os dados) e evitar sobreposições. Assim, a gestão da Cadeia de Custódia Digital (incluindo o controle da coleta, movimentação e acesso aos dados, com a identificação de quem, quando, onde e por que os acessou, bem como quaisquer alterações) é de extrema importância, sendo uma responsabilidade compartilhada por todos os envolvidos na coleta e tratamento de evidências digitais, tendo em vista que esse instituto foi criado justamente para minimizar as falhas probatórias tendo um impacto enorme no âmbito digital, notadamente porque quem apresenta a prova tem o ônus de comprovar sua cronologia, sem o que a prova deverá ser descartada.

A apuração de crimes que se valem do ambiente digital acaba exigindo essa observância de regras, metodologias e procedimentos técnicos. E, nesse sentido, a necessidade de observância da Cadeia de Custódia Digital é reafirmada pela maleabilidade e vulnerabilidade dos dados digitais, principalmente pela ampla possibilidade de criação de materiais altamente convincentes no âmbito probatório.

CONCLUSÃO

Ante todo o exposto, verifica-se que a democratização do acesso à criação de *deepfakes* está cada vez mais atingindo a confiança do público nos tribunais e turvando a busca pela verdade no processo judicial.

No primeiro capítulo, foi feito um breve resumo do papel das provas no processo penal e suas classificações diante da jurisprudência adotada no sistema penal brasileiro à luz do princípio da Busca pela Verdade, bem como a influência da tecnologia na evolução das práticas

de apresentação de provas no sistema jurídico. Baseando-se nos conceitos de provas lícitas e ilícitas, o capítulo conclui que a tecnologia da Inteligência Artificial está transformando o cenário das regras de apreciação de provas e, consequentemente, tornando-as cada vez mais vulneráveis ao erro por essa mesma tecnologia.

No segundo, foi traçado um panorama geral da relação entre o direito penal e a tecnologia. Foi destacada a crescente ocorrência de cibercrimes e delitos digitais nos últimos anos, se discutindo como esses crimes estão causando prejuízos significativos para a sociedade, abrangendo fraudes financeiras, crimes online, invasões de privacidade, pornografia infantil e outras atividades ilegais relacionadas à tecnologia. No que diz respeito à legislação e regulamentação, o capítulo observa que o avanço da tecnologia tem obrigado o Direito a se adaptar para lidar com novos tipos de condutas ilegais facilitadas pela tecnologia. Nesse sentido, a tendência tem sido aplicar as leis penais existentes, mesmo que isso envolva a interpretação analógica para abranger as atividades ilegais relacionadas à tecnologia. Contudo, com a democratização e facilitação do manuseio da IA, foi observado que o sistema jurídico brasileiro não possui nenhuma legislação específica para o controle e limitação dessa tecnologia e seus prejuízos no âmbito penal, tornando, portanto, cada vez mais difícil de acompanhar e lidar com os danos desse instrumento que está em constante evolução.

Por fim, o terceiro capítulo abordou a problemática relacionada à utilização de deepfakes como prova em processos judiciais a partir de casos emblemáticos sobre o assunto, como eles podem afetar o processo penal no que tange à produção probatória e algumas sugestões de como o Estado e a sociedade podem minimizar os possíveis prejuízos do seu uso. Os deepfakes representam uma ameaça significativa no combate à desinformação, pois a tecnologia permite a criação de vídeos falsos altamente convincentes que podem ser usados para manipular a opinião pública, prejudicar reputações e influenciar negativamente a sociedade. Isso pode causar divisões sociais, polarização, disseminação de teorias conspiratórias e uma crise de confiança na autenticidade de conteúdo audiovisual, além de, principalmente, criar material com conteúdo capaz de falsamente incriminar ou retirar a culpa de alguém.

Assim, também foram explorados os crimes relacionados ao uso de deepfakes, destacando a extensa lista de possibilidades, como ameaças, perseguições, difamações e até mesmo acusações criminais baseadas em provas manipuladas. A questão da autenticidade dessas provas torna-se um desafio, especialmente em sistemas jurídicos que aceitam conversas por aplicativos de mensagens como prova.

Além disso, foram analisados os limites e desafios legais, tecnológicos e sociais para combater os deepfakes e possíveis alternativas para prevenir seu uso indevido. A tecnologia envolvida nesse fenômeno está em constante evolução, tornando difícil para os algoritmos de detecção identificar as diferenças entre material autêntico e manipulado. Isso levanta questões sobre como as leis podem acompanhar essas mudanças e garantir a responsabilização por atividades ilegais relacionadas a deepfakes.

Algumas alternativas propostas incluem a autenticação de conteúdo por meio da blockchain, educação pública para identificar manipulações, coleta de dados biométricos para autenticação, implementação de sistemas rigorosos de autenticação de provas audiovisuais, investimento em tecnologias de detecção, inclusão de peritos em tecnologia e análise forense, estabelecimento de uma cadeia de custódia digital e educação jurídica para profissionais da área.

Infelizmente, é notável a pressão do judiciário por resultados rápidos. E essa situação pode acabar abrindo espaço para a aceitação sem questionamento de provas que parecem autênticas quando, na verdade, são falsas. Diante dessa conjuntura, urge-se uma abordagem multidisciplinar para lidar com as ameaças e riscos que os *deepfakes* trazem no contexto judicial. Nesse contexto, verifica-se a necessidade de especialistas em tecnologia, ética e direito colaborar para o desenvolvimento de protocolos robustos de autenticação e avaliação de provas através da análise de sua origem e validação de sua integridade. Como consequência, também se nota uma crescente demanda de ciência dos riscos associados aos *deepfakes* e da necessidade de uma análise crítica das evidências por parte dos magistrados, advogados e partes envolvidas no processo, para que haja menos riscos de fraude e injustiça.

Porém, a partir do que foi observado, é possível observar que o sistema judiciário brasileiro não possui os recursos necessários para garantir a plena autenticidade das provas e, consequentemente, um processo justo e igualitário. Além da limitação dos recursos orçamentários, há também uma limitação da própria tecnologia, visto que é uma área que está em constante desenvolvimento e correção, tornando difícil para os especialistas encontrarem formas de prevenir ou detectar um materil gerado através da Inteligência Artificial. A atual tecnologia de prova pericial não é capaz de verificar, por exemplo, uma montagem fotográfica se ela for bem feixa, “pixel por pixel”, “bit a bit”. Ainda que o art. 159 do CPP preveja a contratação de assistentes técnicos para auxiliar o perito, isso pode refletir em famílias que não possuem condições de contratar tal profissional para verificar a autenticidade da prova. Pode-se dizer também que não há o equilíbrio entre a grande demanda de verificação das provas (quando não forem incontroversas e houver suspeitas de modificação por *deepfakes*) e o número

de profissionais qualificados para essa avaliação que, inclusive, demanda um tempo considerável.

Portanto, enquanto houver essa limitação orcamentária e tecnológica para implementar técnicas de identificação de uso de deepfakes nos processos penais, urge-se um trabalho de capacitação tanto da população quanto dos operadores da área jurídica e tecnológica para a identificação e prevenção da utilização de conteúdo manipulado de forma realista em processos judiciais, tendo em vista que a tecnologia envolvendo inteligência artificial não encontra limites no atual sistema brasileiro.

REFERÊNCIAS

ARBULU, R. App DeepNude, usado para "criar" nudes femininos, é descontinuado. **CanalTech**, 2019. Disponível em: https://canaltech.com.br/internet/app-deepnude-usado-para-criar-nudes-femininos-e-descontinuado-142863/.

BRASIL. Decreto-Lei nº 2.848, de 7 de dezembro de 1940. Institui o Código Penal. **Diário Oficial da União**, Rio de Janeiro, RJ, 31 dez. 1940. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto-lei/del2848compilado.htm

BRASIL. Lei nº 4.737, de 15 de julho de 1965. Institui o Código Eleitoral. **Diário Oficial da União**, Brasília, DF, 16 jul. 1965. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l4737compilado.htm

BRASIL. Lei Federal n.º 12.735, de 30 de novembro de 2012. Dispõe sobre a "responsabilização dos provedores de conexão e de aplicações de internet por danos decorrentes de conteúdo gerado por terceiros" e altera o art. 7º da Lei nº 8.137, de 27 de dezembro de 1990 (Código de Defesa do Consumidor), e o art. 109 da Lei nº 9.279, de 14 de maio de 1996 (Lei da Propriedade Industrial), para os fins que especifica. **Diário Oficial da União**, Brasília, DF, 3 dez. 2012. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12735.htm

BRASIL. Lei Federal n.º 12.737, de 30 de novembro de 2012. Define os crimes cibernéticos e dá outras providências. **Diário Oficial da União**, Brasília, DF, 3 dez. 2012. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/l12737.htm

BRASIL. Lei Federal n.º 12.965, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. **Diário Oficial da União**, Brasília, DF, 24 abr. 2014. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet), e a Lei nº 9.472, de 16 de julho de 1997 (Lei Geral de Telecomunicações). **Diário Oficial da União**, Brasília, DF, 15 ago. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

CAPÍTULO 54 – INTRODUÇÃO ÀS REDES ADVERSÁRIAS GENERATIVAS (GANs – Generative Adversarial Networks). **Deep Learning Book.** Disponível em: https://www.deeplearningbook.com.br/introducao-as-redes-adversarias-generativas-gans-generative-adversarial-networks/. Acesso em: 14 jul 2023.

CRIME ORGANIZADO INTENSIFICARÁ USO DE DEEP FAKES, PREVÊ EUROPOL. **Ciso Advisor,** 2022. Disponível em: https://www.cisoadvisor.com.br/crime-organizado-intensificara-uso-de-deepfakes-preve-europol/. Acesso em 04 ago 2023.

DAMIANI, J. A Voice Deepfake Was Used To Scam A CEO Out Of $243,000. **Forbes**, 2019. Disponível em https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/?sh=1d4d27b62241. Acesso em 06 ago 2023.

DEEPFAKE TIMIT IS A DATABASE OF VIDEOS WHERE FACES ARE SWAPPED USING THE OPEN SOURCE GAN-BASED APPROACH, WHICH, IN TURN, WAS DEVELOPED FROM THE ORIGINAL AUTOENCODER-BASED DEEPFAKE ALGORITHM. **Idiap Research Institute.** 2022. Disponível em: https://www.idiap.ch/en/dataset/deepfaketimit. Acesso em 24 jul 2023.

FBI, ROUBO DE DADOS E DEEPFAKE: Entenda as polêmicas por trás do aplicativo FaceApp. **Expresso PB.** 2022. Disponível em https://expressopb.com.br/fbi-roubo-de-dados-e-deepfake-entenda-as-polemicas-por-tras-do-aplicativo-faceapp/. Acesso em 19 jul 2023.

FILHO, T. **Processo Penal**. V. 1. São Paulo, Ed. Saraiva, 2000, p.41.

FISCHER, M. VFXChris Ume. **The chronicles of Deep Tom Cruise.** Youtube, 2021. Disponível em: https://www.youtube.com/watch?v=nwOywe7xLhs. Acesso em: 28 jul 2023.

GUARNERA, L.; et al. **The Face Deepfake Detection Challenge**. 2022. 8-10. Journal of Imaging. Disponível em: https://www.mdpi.com/2313-433X/8/10/263. Acesso em: 19 jul 2023.

JR., Aury L. **Direito processual penal**. São Paulo: Editora Saraiva, 2023. Disponível em: https://integrada.minhabiblioteca.com.br/#/books/9786553626355/. Acesso em: 12 set. 2023.

MARTINS, C. Deep Nude tira a roupa a fotos de mulheres usando A.I. **Aberto até de madrugada,** 2019. Disponível em: https://abertoatedemadrugada.com/2019/06/deep-nude-tira-roupa-fotos-de-mulheres.html. Acesso em 31 jul 2023.

MOLINA, A, C.; BERENGUEL, O, L. **Deepfake: A evolução das fake news**. Research, Society and Development, v. 11, n. 6, e56211629533, 2022. DOI: http://dx.doi.org/10.33448/rsd-v11i6.29533. Disponível em: https://rsdjournal.org/index.php/rsd/article/view/29533. Acesso em 10 jul 2023.

MORAES, Cristiane Pantoja De. **"Deepfake" como ferramenta de manipulação e disseminação de "fakenews" em formato de vídeo nas redes sociais.** 2020. Biblios, ISSN 1562-4730 No 79 (2020). DOI 10.5195/biblios.2020.864. Disponível em: https://dialnet.unirioja.es/descarga/articulo/8041632.pdf. Acesso em 14 jul 2023.

NUCCI, Guilherme de S. **Manual de Processo Penal**. São Paulo: Grupo GEN, 2022.. Disponível em: https://integrada.minhabiblioteca.com.br/#/books/9786559643691/. Acesso em: 12 set. 2023.

O QUE É REVENGE PORN OU PORNOGRAFIA DE VINGANÇA?. **Favaretto Araújo Abreu Advogados,** 2022. Disponível em: https://favarettoadv.com.br/2023/08/10/o-que-e-revenge-porn-ou-pornografia-de-vinganca/. Acesso em 31 jul 2023.

REDES ADVERSÁRIAS GENERATIVAS. **Explicando IA.** 2022. Disponível em: https://atozofai.withgoogle.com/intl/pt-BR/gans/. Acesso em 19 jul 2023.

ROSA, Alexandre Morais da. **Quando o defensor e a tecnologia viram o jogo no flagrante.** Revista Consultor Jurídico, 2022. Disponível em: https://www.conjur.com.br/2022-nov-04/limite-penal-quando-defensor-tecnologia-viram-jogo-flagrante. Acesso em 13 out 2023.

SCARLETT JONHANSSON FALA SOBRE DEEPFAKE EM VÍDEOS PORNOGRÁFICOS. **Filmow.** 2019. Disponível em: https://filmow.com/noticias/29534/scarlett-johansson-fala-sobre-deepfake-em-videos-pornograficos/. Acesso em: 2 ago 2023.

SILVERMAN, C. Como identificar um "deepfake" como este vídeo do Barack Obama. **BuzzFeed News,** 2018. Disponível em: https://www.buzzfeed.com/br/craigsilverman/como-identificar-deepfake-video-obama-peele. Acesso em: 28 jul 2023.

The Telegraph. **Deepfake video of Volodymyr Zelensky surrendering surfaces on social media**. Youtube, 2022. Disponível em https://www.youtube.com/watch?v=X17yrEV5sl4. Acesso em: 2 ago 2023.

THIES, J. et al. **Face2Face: Real-time Face Capture and Reenactment of RGB Videos.** 2016, pp. 2387-2395. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Disponível em: https://openaccess.thecvf.com/content_cvpr_2016/html/Thies_Face2Face_Real-Time_Face_CVPR_2016_paper.html. Acesso em: 28 jul 2023.

TUCCI, R.L. **Princípios e regras orientadoras do Novo Processo Penal Brasileiro**. São Paulo: Forense, 1986. p. 145.

VELLA, V,. A Bucks County woman created ‘deepfake’ videos to harass rivals on her daughter’s cheerleading squad, DA says. **The Philadelphia Inquirer**, 2021. Disponível em: https://www.inquirer.com/news/bucks-county-raffaela-spone-cyberbullying-deepfake-20210312.html. Acesso em 9 ago 2023.

Welcome A.I. **Using AI deepfake techniques to bring Salvador Dali back to life**. Youtube, 2019. Disponível em: https://www.youtube.com/watch?v=BxlPCLRfk8U. Acesso em: 25 jul 2023.