FACULDADE DE TECNOLOGIA DE SÃO PAULO

Vivian Kimye Akutsu

Deepfake e sua influência em investigações criminais

SÃO PAULO
2025

FACULDADE DE TECNOLOGIA DE SÃO PAULO

Vivian Kimye Akutsu

Deepfake e sua influência em investigações criminais

Trabalho submetido como exigência parcial
para a obtenção do Grau de Tecnólogo em
Análise e Desenvolvimento de Sistemas
Orientador: Maurício Amaral de Almeida

SÃO PAULO
2025

Dedico este trabalho a meus pais Katsuo e Claudete, meu filho Alfredo, a minha família, a meus amigos e meus professores, em especial ao meu orientador Prof. Maurício Almeida do Amaral.

AGRADECIMENTOS

Gratidão é a palavra que resume meu sentimento.

Gratidão à minha família que sempre me apoiou em várias jornadas.

Gratidão aos amigos que compartilharam as aulas, as dores, felicidades e os momentos únicos.

Gratidão a Fatec e seus professores que dividiram experiências, conhecimento e seu tempo.

“O que todos devemos fazer é nos certificar que estamos usando a inteligência artificial de uma maneira que beneficie a humanidade, e não que a deteriore.”

(Tim Cook)

RESUMO

O advento das tecnologias de manipulação de mídia digital, mais especificamente os *deepfakes*, tem gerado discussões sobre seus impactos em diversas áreas, incluindo no campo das investigações criminais. Os *deepfakes* referem-se a vídeos, áudios ou imagens manipuladas por algoritmos de inteligência artificial para criar representações extremamente realistas de pessoas, situações ou eventos que nunca ocorreram. Embora essa tecnologia tenha aplicações em áreas como entretenimento e educação, sua utilização indevida levanta sérias preocupações, quando aplicada a contextos jurídicos e investigativos.

Palavras-Chave: Deepfakes, Inteligência Artificial, Criminal

ABSTRACT

The advent of digital media manipulation technologies, more specifically deepfakes, has generated discussions about their impacts in several areas, including the field of criminal investigations. Deepfakes refer to videos, audios or images manipulated by artificial intelligence algorithms to create extremely realistic representations of people, situations or events that never occurred. Although this technology has applications in areas such as entertainment and education, its misuse raises serious concerns, when applied to legal and investigative contexts.

Keywords: Deepfakes, Artificial Intelligence, Criminal

LISTA DE ILUSTRAÇÕES

Sumário

1 INTRODUÇÃO

As dificuldades na busca pela verdade histórica durante um processo judicial, especialmente no âmbito penal. Destaca-se que a percepção dos fatos é limitada pela cognição humana e pela ausência do juiz na ocorrência direta dos eventos, o que exige que ele se baseie em documentos e testemunhos.

O princípio da imediação é apresentado como uma tentativa de minimizar essas limitações. Ele se aplica essencialmente à fase de julgamento e exige que a decisão judicial seja proferida por quem presenciou a produção das provas e a discussão do caso. Esse princípio valoriza provas mais diretas, como testemunhos presenciais e documentos originais.

Apesar de seu valor, o princípio da imediação não garante a descoberta completa da verdade histórica, funcionando apenas como uma aproximação mediada aos fatos ilícitos, destacando as dificuldades inerentes ao processo judicial na busca pela precisão dos acontecimentos.

2 Inteligência Artificial e o Direito Penal

A rápida evolução das novas tecnologias, especialmente no campo da inteligência artificial (IA) e do aprendizado de máquina, tem introduzido desafios significativos para o Direito contemporâneo. Esses avanços tecnológicos possuem o potencial de transformar profundamente a sociedade, exigindo que o ordenamento jurídico, como instrumento regulador, estabeleça fronteiras claras entre o lícito e o ilícito, o desejável e o indesejável, bem como delimite os riscos permitidos e os proibidos.

No âmbito do direito penal, a emergência e o desenvolvimento de técnicas de IA têm contribuído para a criação de novas formas de condutas criminosas, redefinindo crimes tradicionais e levantando a necessidade de identificar bens jurídicos até então inexistentes na estrutura jurídico-criminal. Por exemplo, a difamação, que historicamente se limitava a declarações verbais ou escritas, agora pode ser perpetrada por meio de conteúdos gerados por IA, como vídeos ou áudios falsificados, conhecidos como deepfakes. Essas mudanças refletem os anseios comunitários por proteção de estados, objetos ou bens de relevância individual ou coletiva. Em outras palavras, o impacto transformador da IA é cada vez mais evidente nos aspectos mais sutis da nossa vida, incluindo aqueles que o direito penal substantivo deve abordar, seja por meio da criação de novas tipificações criminais, da adaptação a modos de execução inéditos ou do reexame de suas categorias dogmáticas e princípios orientadores.

A relação simbiótica entre o direito penal substantivo e o processo penal impõe a este último a necessidade de reflexão e possível adaptação a essa nova realidade. Estudos sobre o impacto de ferramentas algorítmicas de suporte à decisão judicial são comuns, especialmente considerando que seu funcionamento, frequentemente opaco e suscetível a vieses sociais, pode comprometer o direito a um julgamento justo e dificultar o exercício efetivo de garantias de defesa, como o direito ao recurso. Além disso, o uso de IA pode afetar direitos fundamentais como a privacidade. A crescente presença de dispositivos inteligentes, como assistentes virtuais, que registram e armazenam informações do ambiente, pode levar a situações em que a IA atua como

testemunha, fornecendo evidências cruciais em investigações ou processos judiciais, com implicações significativas no âmbito probatório.

Uma questão particularmente relevante é a criação ou manipulação de fotos, vídeos e áudios por meio de IA, conhecidos como deepfakes, e seu impacto no processo penal. A disseminação dessas falsificações sofisticadas representa um desafio para a produção de provas e a busca pela verdade real nos processos judiciais. A capacidade de criar deepfakes como provas em processos judiciais levanta preocupações sérias sobre a integridade do sistema de justiça criminal, pois torna-se difícil distinguir entre provas genuínas e falsificadas. Além disso, a utilização de deepfakes para fins ilícitos, como a difamação ou a disseminação de informações falsas, pode causar danos significativos às vítimas e comprometer a credibilidade das instituições judiciais.

Diante desse cenário, é fundamental que o sistema jurídico desenvolva mecanismos eficazes para identificar e combater o uso indevido de IA no contexto criminal. Isso inclui a implementação de técnicas avançadas de perícia forense digital capazes de detectar manipulações em mídias digitais, bem como a elaboração de legislações específicas que tipifiquem e punam adequadamente condutas relacionadas à criação e disseminação de deepfakes. A criação de legislação específica para abordar crimes relacionados a deepfakes é fundamental, podendo incluir penalidades mais severas para a criação e disseminação indevida dessas falsificações. Além disso, é essencial promover a conscientização sobre os riscos associados ao uso indevido da IA incentivando a educação digital e a alfabetização midiática para que os cidadãos possam identificar e se proteger contra possíveis manipulações.

A integração da IA no direito penal também traz à tona questões éticas e de direitos humanos. A utilização de sistemas automatizados na tomada de decisões judiciais deve ser acompanhada de debates constantes sobre ética, privacidade e direitos fundamentais, assegurando que a tecnologia sirva à justiça, e não o contrário. A transparência nos algoritmos utilizados, a explicabilidade das decisões automatizadas e a possibilidade de revisão humana são aspectos cruciais para garantir a legitimidade e a justiça no uso da IA no sistema penal.

Em suma, a evolução da inteligência artificial apresenta desafios complexos para o direito penal, exigindo adaptações tanto no âmbito substantivo quanto processual. A criação de novas tipificações criminais, a adaptação de procedimentos processuais e o desenvolvimento de mecanismos eficazes de detecção e prevenção de abusos são medidas essenciais para assegurar que o sistema jurídico acompanhe as transformações tecnológicas, protegendo os direitos fundamentais e mantendo a integridade da justiça.

3 Geração de Falsificações

A seguir, são apresentados os tipos mais comuns de falsificações, assim como os métodos mais utilizados para a geração das mesmas.

3.1.1 Tipos de Falsificações

As falsificações digitais podem ser classificadas, conforme o grau de manipulação empregado, em quatro categorias principais (DANG et al., 2019). TOLOSANA et al. (2016) fornece uma descrição detalhada de cada uma delas, que são apresentadas a seguir de forma sucinta, organizadas em ordem crescente de complexidade na manipulação.

3.1.2 Expressões faciais

Trata-se da alteração das expressões faciais de um indivíduo, sendo essa modificação realizada por meio da transferência de expressões de uma pessoa para outra (AVERBUCH-ELOR et al., 2017). Esse tipo de manipulação pode ser subdividido em duas categorias: *source-to-target*, em que as expressões faciais de uma pessoa são transferidas para outra, e *self-reenactment*, em que a mesma imagem atua como

fonte e destino da manipulação. A técnica mais amplamente utilizada para esse tipo de modificação é denominada *Face2Face* (THIES et al., 2016).
A Figura 1 ilustra esse tipo de manipulação.

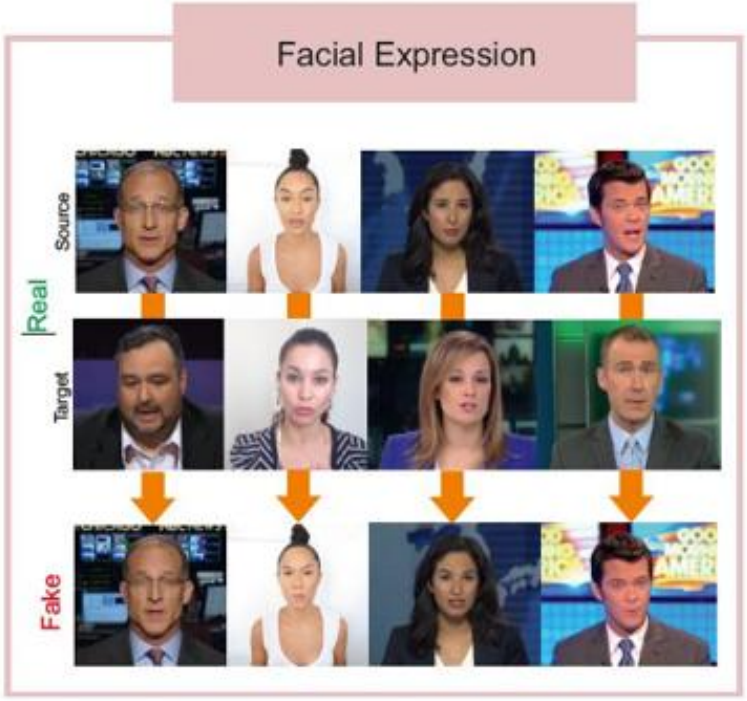

Figura 1 - Manipulação de expressões faciais. Fonte: (TOLOSANA et al., 2016).

3.1.3 Atributos faciais

Refere-se à modificação de atributos faciais, como cor da pele ou do cabelo, gênero, idade, ou ainda à adição de elementos como óculos. Esse tipo de manipulação é geralmente realizado por meio de Redes Geradoras Adversárias (GANs), conforme descrito na Seção 2.2.5 (CHOI et al., 2018). Um exemplo conhecido de aplicação que utiliza esse tipo de técnica é o aplicativo FaceApp (FaceApp, 2017).

A Figura 2 ilustra esse tipo de manipulação.

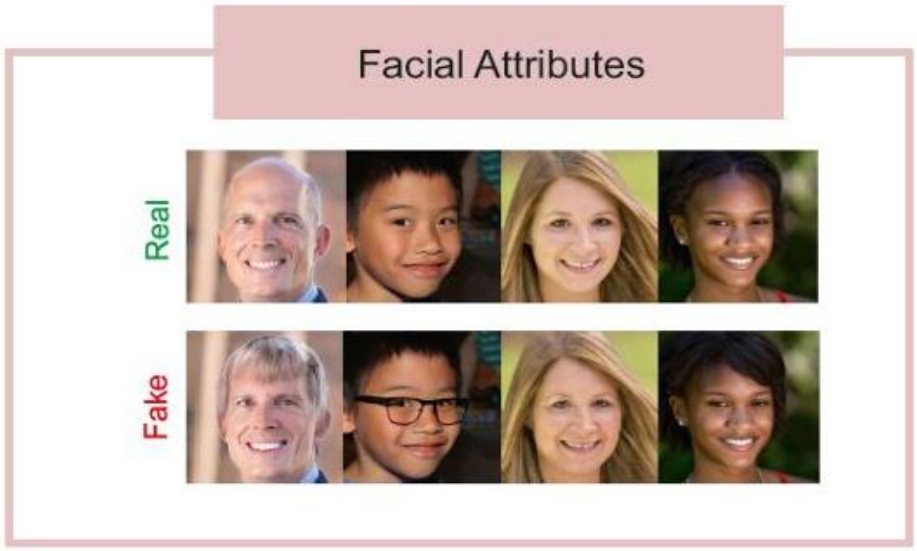

Figura 2 - Manipulação de atributos faciais. Fonte: (TOLOSANA et al., 2016).

3.1.4 Troca de identidades (Face Swap)

Este é, possivelmente, o tipo de manipulação mais amplamente reconhecido, por ter sido o precursor das pesquisas atuais envolvendo deepfakes e por sua ampla disseminação em plataformas como o YouTube. Um exemplo notável é o canal do brasileiro Bruno Sartori, que utiliza esse tipo de manipulação para criar sátiras políticas e se autodenomina um “deepfaker” (SARTORI, 2012). Essa técnica consiste na substituição do rosto de uma pessoa pelo de outra. Para sua realização, podem ser empregadas duas abordagens principais: (i) a abordagem clássica, baseada em técnicas de computação gráfica, como o método FaceSwap (KOWALSKI, 2018), e (ii) a abordagem mais recente, que utiliza técnicas de deep learning, conhecida como Deepfakes. Um exemplo de aplicação comercial que emprega essa tecnologia é o aplicativo ZAO (LOUBAK, 2019).

A Figura 3 apresenta exemplos desse tipo de manipulação.

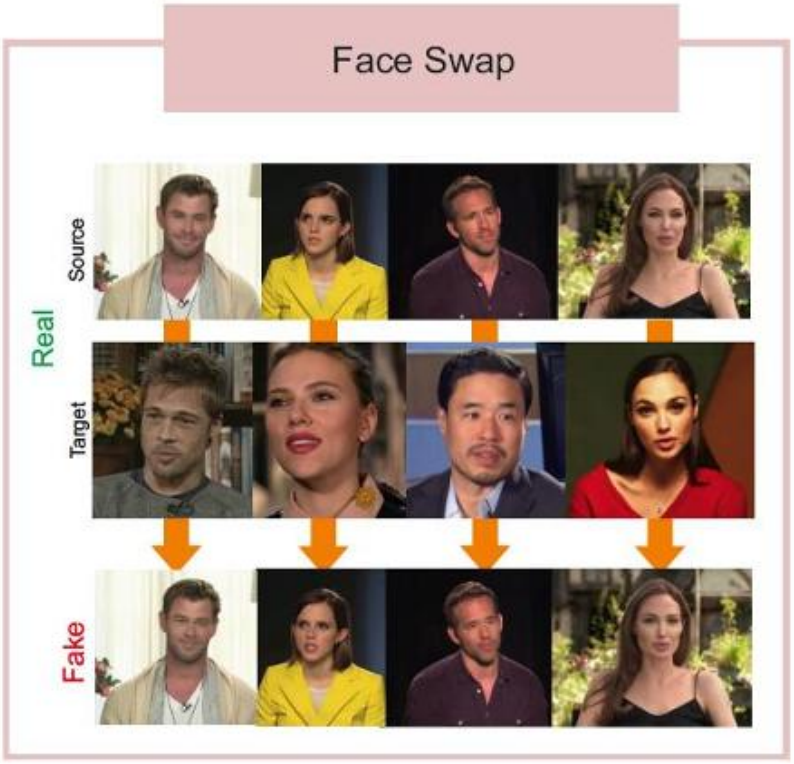

Figura 3 - Manipulação do tipo troca de identidades. Fonte: (TOLOSANA et al., 2016).

3.1.5 Síntese facial

Esse tipo de manipulação é responsável pela criação de rostos completamente novos, ou seja, pela síntese de faces inexistentes na realidade. Em geral, esse processo é realizado por meio do uso de Redes Geradoras Adversárias (GANs) (KARRAS; LAINE; ALLA, 2019). Os resultados obtidos apresentam um nível impressionante de qualidade e realismo (WEST; BERGSTROM, 2019; WANG, 2019). A Figura 4 ilustra exemplos desse tipo de manipulação.

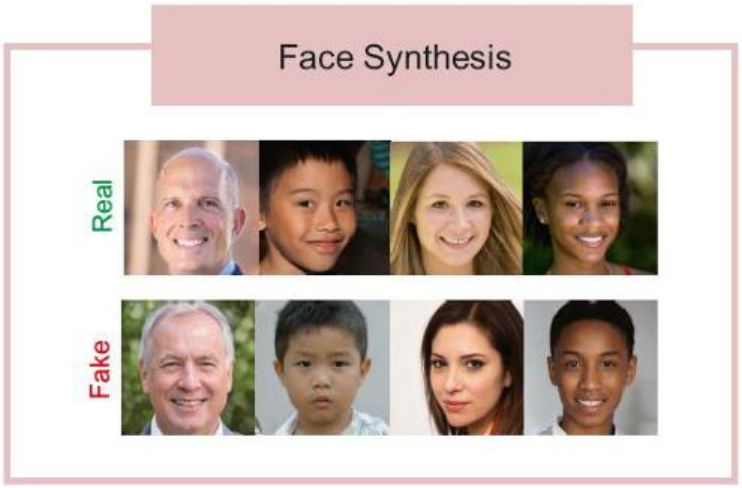

Figura 4 - Manipulação do tipo síntese facial. Fonte: (TOLOSANA et al., 2016).

4 DeepFakes

O modelo inicialmente adotado baseava-se em uma arquitetura de autoencoder-decoder. Nessa configuração, o autoencoder é responsável por extrair características latentes das imagens faciais representações comprimidas e de baixa resolução, enquanto o decoder reconstrói as imagens a partir dessas representações. Para que a substituição de rostos seja realizada, utilizam-se dois decoders, cada um treinado com uma base de dados distinta (rostos de A e rostos de B), os quais compartilham um mesmo encoder. Essa estratégia, conforme descrito por NGUYEN et al. (2019), permite que o encoder aprenda padrões comuns entre os conjuntos de imagens, como a forma dos olhos, nariz e a posição da boca. Por fim, realiza-se a troca dos decoders, de modo que, a partir das características extraídas de uma face A, seja possível gerar uma nova imagem com as características faciais correspondentes à identidade de B. Essa abordagem é empregada em diversos projetos, como o DeepFaceLab (PEROV, 2018).

A Figura 5 ilustra o processo descrito.

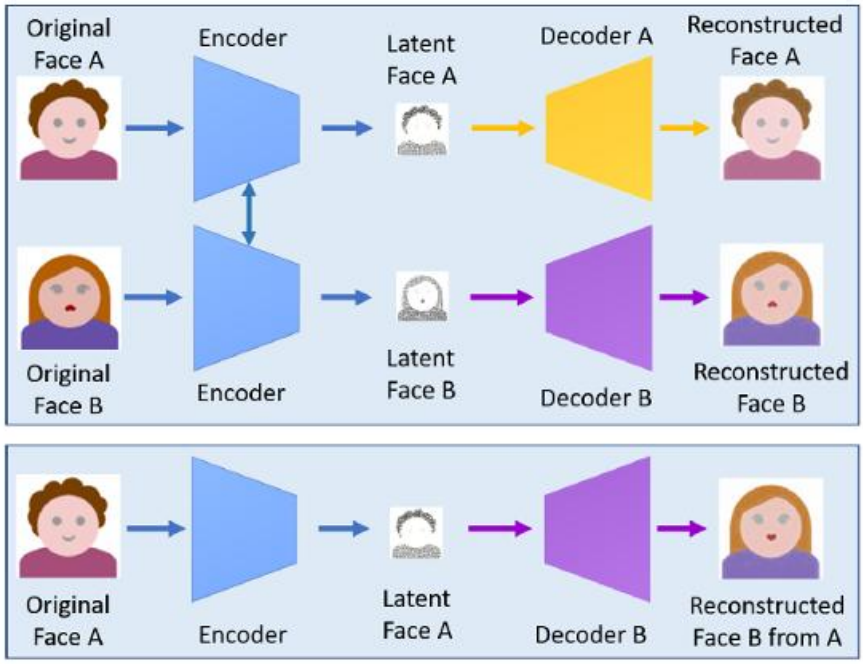

Figura 5 - O método Deepfakes. Fonte: (NGUYEN et al., 2019).

4.1 Face2Face

A técnica denominada Face2Face refere-se a um sistema de manipulação facial que transfere expressões faciais de um vídeo de origem para um vídeo de destino, preservando a identidade da pessoa alvo. Sua implementação baseia-se na utilização de dois vídeos como entrada, com a seleção manual de frames-chave. Esses frames são empregados na geração de uma reconstrução densa da face, a qual permite a re-sintetizar da expressão facial com variações tanto nas expressões quanto nas condições de iluminação.

4.2 NeuralTextures

Essa técnica tem como objetivo a manipulação de expressões faciais. Para isso, utiliza os dados do vídeo original a fim de aprender a chamada “textura neural” da

pessoa alvo, a qual é manipulada e, posteriormente, recombinada por meio de uma rede de renderização.

5 Deepfake e o Direito Probatório

A manipulação de imagens, sejam elas estáticas ou em movimento, e de áudios não é uma prática exclusiva do século XXI. Sua origem remonta aos primórdios das técnicas de gravação e captura, entrelaçando-se com a própria história da fotografia, do vídeo e da gravação de som, mesmo antes da era digital.

Um exemplo notável dessa prática ocorreu em 1840, quando Hippolyte Bayard, pioneiro da fotografia, produziu o "Autorretrato como Homem Afogado". Nesta imagem, Bayard aparece como um cadáver, supostamente vítima de suicídio por afogamento. Essa fotografia foi uma forma de protesto contra o reconhecimento insuficiente de sua contribuição para a invenção da fotografia, especialmente em comparação ao destaque recebido por Louis Daguerre. No verso da imagem, Bayard escreveu uma nota sarcástica explicando sua decisão fictícia de tirar a própria vida devido à negligência das autoridades e da sociedade em relação ao seu trabalho.

Ainda no século XIX, surgiram as chamadas "fotografias espíritas", que utilizavam técnicas de dupla exposição para sobrepor imagens, criando a ilusão de aparições fantasmagóricas. Essas imagens eram frequentemente apresentadas como evidências de fenômenos sobrenaturais, explorando a credulidade popular da época. Com o avanço das técnicas fotográficas, a manipulação de imagens tornou-se mais sofisticada e acessível. Apesar de sua suscetibilidade a alterações, a fotografia passou a ser aceita como meio de prova nos tribunais, sujeita à avaliação crítica dos julgadores. No Reino Unido, por exemplo, a admissibilidade de material fotográfico como prova incriminatória é reconhecida em determinadas circunstâncias, como:
Quando a imagem é suficientemente nítida, permitindo que o júri a compare diretamente com o réu presente no tribunal.

Quando uma testemunha conhece bem o réu e pode identificá-lo na fotografia apresentada.

Quando uma testemunha, mesmo sem conhecer o réu, analisa extensivamente as imagens e adquire conhecimento especializado que o júri não possui, podendo então fornecer uma identificação baseada na comparação entre as imagens e uma fotografia recente do réu.

Quando um especialista qualificado em mapeamento facial oferece testemunho baseado na comparação entre as imagens da cena e uma fotografia atual do réu, desde que ambas estejam disponíveis para o júri.

Essas diretrizes refletem a cautela necessária na avaliação de provas fotográficas, reconhecendo tanto seu valor potencial quanto suas limitações.

Com o advento da inteligência artificial (IA), a criação e manipulação de imagens e áudios atingiram um novo patamar. A IA permite a geração de conteúdos falsificados extremamente realistas, conhecidos como "deepfakes", que podem ser utilizados para fins diversos, desde entretenimento até fraudes e desinformação. Essa evolução tecnológica apresenta desafios significativos para o direito penal e o processo judicial, especialmente no que tange à autenticidade das provas e à proteção dos direitos individuais.

Portanto, embora a manipulação de mídias não seja uma novidade, as ferramentas modernas potencializam seus efeitos e complexificam sua detecção, exigindo uma constante adaptação das normas jurídicas e dos métodos de análise forense para assegurar a integridade e a justiça nos processos legais.

A inteligência artificial (IA) introduz inovações significativas na criação e manipulação de imagens e áudios, especialmente através de técnicas como o deep learning. O deep learning, ou aprendizado profundo, é uma subcategoria do aprendizado de máquina que utiliza redes neurais artificiais para modelar padrões complexos em dados. Essa abordagem tem sido aplicada em diversas áreas, incluindo reconhecimento de voz, processamento de linguagem natural e descoberta de medicamentos, destacando-se por sua capacidade de aprender diretamente dos dados sem supervisão humana intensiva.

Uma característica distintiva do deep learning é sua habilidade de operar de forma autônoma, enquadrando-se na categoria de "aprendizado não supervisionado". Diferentemente do aprendizado supervisionado, que requer a intervenção humana para fornecer conjuntos de dados rotulados, o aprendizado não supervisionado

permite que os algoritmos identifiquem padrões e estruturas nos dados por conta própria. Essa autonomia reduz a necessidade de grandes volumes de dados rotulados e minimiza as limitações associadas à intervenção humana, tornando o processo mais eficiente e escalável.

Os deepfakes, produtos dessas técnicas avançadas de IA, podem ser classificados em cinco categorias principais: substituição facial, reencenação facial, geração de rostos, síntese de fala e falsificações superficiais (shallowfakes). A substituição facial envolve a transferência da face de uma pessoa para outra em um vídeo ou imagem, criando a ilusão de que alguém está realizando ações que nunca ocorreram. A reencenação facial permite manipular expressões e movimentos faciais de uma pessoa, sincronizando-os com outro áudio ou vídeo, o que pode ser utilizado para falsificar declarações ou emoções. A geração de rostos utiliza IA para criar imagens realistas de pessoas que não existem, enquanto a síntese de fala envolve a criação ou modificação de discursos, replicando a voz de indivíduos específicos. As falsificações superficiais referem-se a edições mais simples, utilizando técnicas básicas de edição para alterar conteúdos existentes.

Em 2022, ferramentas de conversão de texto para imagem, como DALL-E 2, Stable Diffusion e Midjourney, tornaram-se amplamente acessíveis ao público. Essas plataformas permitem que usuários descrevam textualmente a imagem desejada, e a IA gera uma representação visual correspondente, ampliando as possibilidades criativas e levantando questões sobre autenticidade e direitos autorais. Recursos adicionais, como inpainting (edição de partes específicas de uma imagem), outpainting (expansão de uma imagem além de suas bordas originais) e image-to-image (transformação de uma imagem baseada em uma descrição textual), oferecem ferramentas poderosas para a criação e manipulação de conteúdo visual.

Diante dessas inovações, surge uma preocupação crescente sobre a confiança na autenticidade do que vemos e ouvimos. Conforme alertado por organizações como a Europol, Nações Unidas e Trend Micro, conteúdos deepfake podem ser maliciosamente apresentados como evidências legítimas, comprometendo investigações criminais e processos judiciais. Essa realidade impõe desafios significativos ao sistema de justiça, exigindo novas abordagens legais e tecnológicas para identificar e mitigar os riscos associados às manipulações digitais avançadas.

No contexto jurídico brasileiro, embora existam leis como o Marco Civil da Internet (Lei nº 12.965/2014) e a Lei Geral de Proteção de Dados (Lei nº 13.709/2018) que abordam aspectos relacionados ao uso da internet e proteção de dados, ainda não há uma tipificação penal específica para deepfakes. Essa lacuna legal pode resultar em dificuldades para responsabilizar criminalmente os criadores e disseminadores desse tipo de conteúdo, especialmente quando utilizado para adulteração de provas, disseminação de notícias falsas ou pornografia não consensual.

Em suma, a evolução da IA e das técnicas de deep learning trouxe ferramentas poderosas para a criação e manipulação de conteúdo digital. Embora essas tecnologias ofereçam oportunidades criativas e avanços em diversas áreas, também apresentam desafios significativos para a autenticidade da informação e a integridade dos processos judiciais. É imperativo que o sistema jurídico se adapte a essas novas realidades, desenvolvendo mecanismos legais e técnicos para identificar e combater o uso malicioso de deepfakes e outras formas de manipulação digital.

Uma resposta eficaz aos riscos apresentados pelos deepfakes exige uma abordagem multidisciplinar. Conforme observado, "o que o digital quebra ele também pode reparar, não muito diferente da luta interminável entre vírus e antivírus. No nosso caso, além de educarmos as pessoas, adquirir novas sensibilidades e ter o enquadramento legal adequado". É fundamental promover a educação e a literacia digital, garantindo que o acesso às novas tecnologias seja acompanhado por uma atitude crítica e reflexiva em relação às informações consumidas. Preparar a sociedade para a possível má utilização de deepfakes é essencial, mas a própria tecnologia também pode ser aliada na mitigação desse problema, por meio de ferramentas automatizadas de detecção de deepfakes ou da verificação da origem de conteúdos audiovisuais utilizando blockchains.

No âmbito do direito processual penal, é inegável que os deepfakes podem comprometer o conjunto probatório e influenciar a decisão final do julgador. Imagine-se, por exemplo, a possibilidade de serem apresentadas como provas vídeos ou fotos adulterados ou criados artificialmente. No contexto normativo português atual, destaca-se o princípio da livre apreciação da prova. Estabelecido no artigo 127.º do Código de Processo Penal (CPP), este princípio determina que "a prova é apreciada

segundo as regras da experiência e a livre convicção da entidade competente". Isso implica que o juiz deve valorar a prova não de forma predeterminada pela lei, mas de acordo com as regras da experiência e sua livre convicção. Contudo, este princípio não é absoluto, apresentando exceções como as regras relativas ao depoimento indireto (artigo 129.º do CPP) ou às vozes públicas e convicções pessoais (artigo 130.º do CPP). De modo geral, aplica-se sem limitações especiais à apreciação de reproduções fotográficas, cinematográficas ou fonográficas.

Diante do surgimento dos deepfakes, surge a necessidade de reavaliar os princípios que orientam o direito processual penal português e, consequentemente, o direito positivo. Manter o status quo normativo – de sujeição à livre apreciação da prova – apresenta o inconveniente do relativo, senão generalizado, desconhecimento dessa nova realidade. Como pode ser efetivado o princípio da livre apreciação da prova nos dias atuais se a compreensão de quem procede à valoração estiver baseada em pressupostos desatualizados, próprios de quem confunde passado, presente e futuro, e desconhece os perigos atuais para a sociedade e para a administração da justiça em particular? Sem uma robusta literacia digital que acompanhe o domínio da criação, interpretação e aplicação do direito, corre-se o risco de um fenômeno de obsolescência material do direito.

A crescente sofisticação dos deepfakes apresenta desafios significativos para o sistema de justiça penal, especialmente no que tange à admissibilidade e autenticidade de provas digitais. Uma abordagem para mitigar a manipulação do processo judicial por meio de deepfakes seria estabelecer critérios rigorosos para a aceitação de evidências fotográficas, cinematográficas ou fonográficas. Isso implicaria atribuir ao responsável pela apresentação da prova o ônus de demonstrar sua autenticidade, possivelmente por meio de um laudo pericial que ateste a ausência de indícios de adulteração ou criação artificial, conforme os conhecimentos tecnológicos atuais.Entretanto, essa medida suscita preocupações. A exigência de um laudo pericial pode sobrecarregar financeiramente as partes envolvidas, especialmente se os custos forem arcados por quem apresenta a prova. Isso poderia desencorajar o uso de evidências digitais legítimas, empobrecendo o conjunto probatório disponível e potencialmente comprometendo a busca pela verdade no processo penal.

Paralelamente, a produção de prova pericial, conforme os artigos 151 e seguintes do Código de Processo Penal (CPP), surge como uma resposta direta aos desafios impostos pelos deepfakes. Dada a complexidade técnica envolvida na identificação de conteúdos sintéticos, a intervenção de peritos qualificados é essencial. Contudo, para que o magistrado reconheça a necessidade de tal perícia, é fundamental que esteja ciente das possibilidades de falsificação proporcionadas pelas tecnologias atuais. A falta de familiaridade com esses avanços pode resultar na não solicitação de perícias essenciais, comprometendo a integridade do julgamento.

A jurisprudência brasileira já reconhece a importância da cadeia de custódia na admissibilidade de provas digitais. A Quinta Turma do Superior Tribunal de Justiça (STJ) decidiu que provas obtidas de dispositivos móveis são inadmissíveis se não forem adotados procedimentos que garantam a idoneidade e integridade dos dados extraídos. Essa decisão destaca a necessidade de rigor no tratamento de provas digitais, dada sua susceptibilidade a alterações imperceptíveis.

Diante desse cenário, é imperativo que os operadores do direito sejam capacitados para compreender as nuances das novas tecnologias e seus impactos no processo penal. Somente com uma sólida literacia digital será possível assegurar que a justiça não seja comprometida por evidências manipuladas, protegendo, assim, os direitos fundamentais dos envolvidos.

6 Pontos a favor do uso da tecnologia Deepfake

6.1 Aplicações legítimas

A tecnologia deepfake, quando utilizada de forma ética e dentro dos limites legais, oferece uma ampla gama de aplicações legítimas que beneficiam diversas áreas. No setor de entretenimento, por exemplo, é possível recriar digitalmente atores falecidos para que possam "participar" de novos filmes, garantindo continuidade narrativa em franquias ou projetos cinematográficos. Na área educacional, deepfakes podem ser utilizados para simulações realistas em treinamentos, como nos cursos de medicina, onde estudantes podem interagir com pacientes virtuais que simulam sintomas complexos. O uso de deepfakes para recriar figuras históricas para fins didáticos também amplia o potencial educativo, oferecendo uma experiência visualmente mais impactante e interativa para os alunos. No campo do marketing e da publicidade, deepfakes também permitem a criação de anúncios personalizados com influenciadores ou celebridades, expandindo a comunicação de marcas de maneira inovadora.

6.2 Avanços tecnológicos

Ferramentas avançadas como DALL-E 2, Stable Diffusion e Midjourney representam uma evolução significativa nas técnicas de criação e manipulação de imagens. Essas plataformas, baseadas em inteligência artificial, permitem que usuários descrevam em texto a imagem desejada, gerando resultados visuais incrivelmente realistas. Além do potencial criativo, essas ferramentas podem contribuir para projetos acadêmicos, design gráfico e até para a criação de ambientes virtuais realistas em videogames ou simulações. O avanço dessas tecnologias amplia as possibilidades criativas, oferecendo soluções inovadoras para indústrias que dependem de recursos visuais, ao mesmo tempo que estimulam discussões sobre ética e direitos autorais.

6.3 Auxílio na investigação criminal

A tecnologia deepfake também pode ser uma poderosa ferramenta na área criminal quando empregada para fins investigativos. Técnicas avançadas de reconstrução facial, baseadas em IA, permitem recriar com maior precisão as características físicas de suspeitos ou vítimas. Essa tecnologia pode ser usada para reconstituir cenas de crimes, facilitando o trabalho das autoridades na busca por evidências visuais. Em situações onde as imagens originais estão desfocadas ou incompletas, a IA pode ajudar a melhorar a nitidez e os detalhes para que sejam analisados com mais precisão. Isso contribui para identificar suspeitos ou mesmo reconstruir eventos em que não havia registros claros.

6.4 Detecção de fraudes

Embora os deepfakes sejam conhecidos por facilitar manipulações visuais e sonoras, a própria inteligência artificial pode ser utilizada para identificar essas adulterações. Algoritmos específicos de detecção de deepfakes estão em constante evolução, permitindo a identificação de inconsistências em imagens e vídeos, como piscadas irregulares, movimentos labiais incoerentes ou alterações anômalas na textura da pele. Essa tecnologia de detecção é fundamental para assegurar a integridade de provas digitais e combater fraudes que possam comprometer processos judiciais, investigações criminais e a reputação de indivíduos ou empresas.

7 Pontos contra e preocupações destacadas

7.1 Manipulação criminosa

O uso indevido de deepfakes apresenta sérios riscos à segurança social e jurídica. Criminosos podem empregar essa tecnologia para manipular vídeos, áudios e imagens de forma extremamente realista, criando provas falsas que comprometam a busca pela verdade em processos judiciais. Essa manipulação pode envolver a

falsificação de declarações de autoridades, imputando a pessoas inocentes ações que jamais cometeram. Além disso, em cenários políticos, deepfakes podem ser utilizados para espalhar desinformação e manipular a opinião pública, causando impactos significativos na estabilidade democrática.

7.2 Impacto na credibilidade da justiça

O avanço dos deepfakes apresenta um risco crescente para o sistema judiciário, uma vez que a apresentação de provas manipuladas pode distorcer os fatos e influenciar injustamente a decisão do magistrado. Essa incerteza sobre a autenticidade de provas digitais afeta a confiança da sociedade na credibilidade da justiça, tornando mais difícil distinguir entre evidências reais e manipuladas. Para preservar essa confiança, torna-se imprescindível que os profissionais do direito desenvolvam habilidades técnicas para reconhecer e contestar possíveis fraudes baseadas em deepfakes.

7.3 Riscos sociais e individuais

O impacto dos deepfakes se estende para além do universo jurídico, afetando também a esfera pessoal e social. A criação de conteúdos pornográficos não consensuais, utilizando a imagem de indivíduos sem autorização, é uma das aplicações mais danosas dessa tecnologia. Além disso, deepfakes podem ser empregados para disseminar informações falsas, promovendo difamação e prejudicando a reputação de figuras públicas e cidadãos comuns. Essas manipulações geram danos emocionais e sociais consideráveis, com potencial para provocar crises familiares, profissionais e institucionais.

7.4 Lacuna legal

Embora a legislação brasileira contemple normas relacionadas à proteção digital, como o Marco Civil da Internet e a Lei Geral de Proteção de Dados (LGPD),

ainda não existe uma tipificação penal específica para o uso indevido de deepfakes. Essa ausência de legislação específica dificulta a responsabilização de indivíduos que produzem e disseminam esses conteúdos fraudulentos, especialmente em casos que envolvem falsificação de provas ou manipulações destinadas a prejudicar terceiros. A criação de um arcabouço legal específico é essencial para coibir essas práticas e proteger vítimas desse tipo de crime.

7.5 Desafios na detecção

A identificação de deepfakes exige técnicas sofisticadas de perícia forense digital, capazes de analisar detalhes minuciosos que revelam indícios de manipulação. Entretanto, esse processo é altamente especializado e pode gerar custos elevados, tornando-se inacessível para algumas partes envolvidas em processos judiciais. Além disso, a constante evolução da tecnologia deepfake torna a detecção ainda mais desafiadora, exigindo que especialistas estejam sempre atualizados para acompanhar as novas técnicas de manipulação digital.

8 Deepfakes e suas Implicações Jurídicas

8.1 Causas e Efeitos

A rápida evolução da inteligência artificial (IA), especialmente com o avanço das técnicas de aprendizado profundo (deep learning), tem proporcionado avanços impressionantes na criação e manipulação de conteúdos digitais, especialmente imagens e áudios. Essa revolução tecnológica ampliou significativamente as possibilidades criativas e comunicativas, impactando áreas como o entretenimento, a educação e a publicidade.

Por meio de redes neurais artificiais complexas, as técnicas de deep learning permitem que máquinas aprendam padrões visuais e sonoros com alto nível de precisão, possibilitando a criação de deepfakes extremamente realistas. Essa inovação foi inicialmente explorada de forma benéfica, como na recriação de

personagens históricos para fins educacionais ou na indústria cinematográfica para dar vida a atores falecidos.

Contudo, essa mesma capacidade de gerar conteúdos realistas também despertou preocupações alarmantes, especialmente no contexto jurídico. A produção de deepfakes com finalidades criminosas introduziu uma nova camada de complexidade ao sistema judicial. A adulteração de imagens, vídeos e áudios pode manipular provas, distorcer depoimentos e influenciar a percepção dos julgadores, prejudicando a busca pela verdade real no processo penal.

Além disso, o uso de deepfakes tem sido explorado em práticas ilícitas como difamações, chantagens, fraudes e manipulação política. Esse cenário representa uma grave ameaça à integridade do sistema judicial e à segurança jurídica, uma vez que a confiança na autenticidade das provas é fundamental para a tomada de decisões justas e precisas. Assim, a disseminação dessa tecnologia sem o devido controle expõe não apenas indivíduos, mas também instituições à insegurança e ao descrédito.

8.2 Dedutibilidade a Partir dos Experimentos

Diversos experimentos têm revelado as possibilidades e limitações da tecnologia de deepfakes. Ferramentas como DALL-E 2, Stable Diffusion e Midjourney ilustram a sofisticação alcançada pelas técnicas de inteligência artificial na geração de imagens hiper-realistas a partir de descrições textuais. Essas plataformas permitem criar representações visuais extremamente detalhadas, possibilitando desde ilustrações criativas até manipulações visuais que podem se passar por registros reais.

No entanto, esses avanços também impulsionaram o desenvolvimento de ferramentas especializadas na detecção de manipulações digitais. Pesquisas apontam que algoritmos projetados para identificar deepfakes conseguem detectar padrões e anomalias que não são perceptíveis ao olho humano. Entre os principais indícios detectáveis destacam-se:

8.2.1 Movimentos oculares incoerentes

Muitos deepfakes apresentam piscadas incomuns ou comportamento anômalo das pálpebras, visto que a IA ainda tem dificuldade em replicar esses movimentos de forma natural.

8.2.2 Falhas na iluminação e sombras

A luz pode refletir de maneira desigual em diferentes áreas do rosto ou criar sombras incompatíveis com o ambiente.
Inconsistências nas expressões faciais: Pequenas variações na musculatura do rosto, que são naturais em seres humanos, tendem a ser reproduzidas de forma imperfeita pelos algoritmos.

Esses experimentos revelam que, embora os deepfakes sejam altamente sofisticados, ainda existem fragilidades que permitem sua identificação por meio de ferramentas forenses especializadas. Entretanto, a evolução tecnológica é constante e exige que esses mecanismos de detecção acompanhem o desenvolvimento das técnicas de manipulação digital.

9 Contradições e Desafios

A tecnologia dos deepfakes apresenta uma dualidade significativa, gerando tanto benefícios quanto riscos. Quando utilizada eticamente, essa tecnologia pode contribuir para o aprimoramento de investigações criminais, reconstituindo cenas de crime, recriando rostos de suspeitos com base em descrições ou restaurando imagens danificadas. No entanto, o potencial destrutivo dos deepfakes é igualmente expressivo.

O principal desafio reside na dificuldade de distinguir conteúdos manipulados de registros autênticos, especialmente quando são utilizados de forma criminosa para manipular depoimentos, adulterar provas ou promover desinformação. Essa problemática é agravada pela ausência de uma legislação penal específica que tipifique e regulamente o uso ilícito dos deepfakes. Essa lacuna jurídica cria um ambiente de impunidade, dificultando a responsabilização de infratores e aumentando os riscos à segurança digital e à justiça.

O paradoxo enfrentado pelo sistema jurídico é evidente: ao mesmo tempo que a tecnologia oferece meios inovadores para aprimorar a coleta de provas e facilitar investigações, ela também introduz elementos que podem fragilizar a confiança no processo penal.

10 Aplicação Prática dos Estudos

Diante do potencial danoso dos deepfakes, é imprescindível que os avanços tecnológicos sejam incorporados aos procedimentos jurídicos, especialmente na investigação criminal e na avaliação de provas.

As ferramentas de detecção de deepfakes devem ser amplamente adotadas pelos órgãos responsáveis pela produção e validação de provas digitais. Algoritmos forenses especializados, desenvolvidos para identificar manipulações visuais e sonoras, são essenciais para assegurar a veracidade de conteúdos audiovisuais apresentados em processos judiciais.

Contudo, apenas a utilização dessas ferramentas não é suficiente. É igualmente necessário que os operadores do direito, incluindo magistrados, promotores e advogados, sejam capacitados para compreender as características técnicas dos deepfakes e suas implicações legais.

Além disso, o respeito à cadeia de custódia digital deve ser rigoroso, garantindo que os elementos probatórios permaneçam íntegros e livres de manipulações desde sua obtenção até sua apresentação em juízo. Essa prática é fundamental para evitar que deepfakes sejam incorporados ao conjunto probatório sem a devida verificação técnica.

11 Elaboração de Teoria

Com base nas ameaças e oportunidades que os deepfakes apresentam, proponho uma teoria que combine aspectos jurídicos e tecnológicos para mitigar os riscos dessa tecnologia. Essa teoria deve fundamentar-se nos seguintes pilares:

11.1 Tipificação Penal Específica

É essencial que a legislação brasileira evolua para incluir a criação, disseminação e uso malicioso de deepfakes como condutas tipificadas no Código Penal. Essa medida proporcionaria maior segurança jurídica e facilitaria a responsabilização criminal dos infratores.

11.2 Protocolos de Admissibilidade de Provas Digitais

Para reduzir os riscos de manipulação probatória, sugere-se a exigência de laudos periciais específicos que atestem a autenticidade de conteúdos digitais apresentados como provas nos tribunais.

11.3 Centros Especializados na Detecção de Deepfakes

A criação de núcleos técnicos compostos por profissionais da área jurídica e da computação poderia fornecer suporte especializado na identificação e análise de conteúdos suspeitos, contribuindo para decisões judiciais mais seguras e fundamentadas.

Essa teoria visa estabelecer um equilíbrio entre o potencial inovador da inteligência artificial e a proteção da segurança jurídica e dos direitos fundamentais.

12 Sugestão de Pesquisa Complementar

Embora os avanços tecnológicos na identificação de deepfakes sejam significativos, ainda há desafios a serem superados. Como continuidade desta pesquisa, sugiro um estudo focado no desenvolvimento de algoritmos que não apenas detectem deepfakes, mas também identifiquem sua origem e intenção.

Essa abordagem permitiria distinguir manipulações feitas para fins lícitos (como em produções cinematográficas ou didáticas) de deepfakes criados com objetivos criminosos. Além disso, ferramentas que rastreiem a origem dos conteúdos manipulados podem auxiliar na identificação dos autores dessas falsificações, facilitando investigações e responsabilizações penais.

A criação de padrões internacionais para a autenticação de conteúdos audiovisuais também se mostra uma área promissora de pesquisa, visando garantir que materiais legítimos possam ser facilmente reconhecidos, enquanto deepfakes maliciosos sejam prontamente identificados e descartados.

Conclusão

O presente trabalho teve como objetivo analisar os impactos da tecnologia de deepfakes na sociedade contemporânea, com ênfase nas suas implicações jurídicas e criminais. Verificou-se que, embora essa tecnologia apresente aplicações legítimas em áreas como o entretenimento, a educação e a comunicação, seu uso indevido representa uma ameaça concreta à segurança jurídica, à autenticidade das provas e à confiança nas instituições públicas.

No campo do Direito Penal, os deepfakes possibilitam a prática de novas condutas criminosas, como falsificação de provas, difamação, fraude e manipulação da opinião pública. Contudo, a inexistência de legislação específica no ordenamento jurídico brasileiro dificulta a responsabilização penal eficaz dos indivíduos que utilizam essa tecnologia para fins ilícitos, evidenciando lacunas normativas que necessitam ser preenchidas com urgência.

A confiabilidade das provas digitais constitui uma das principais preocupações no contexto jurídico, uma vez que conteúdos audiovisuais manipulados podem comprometer a integridade do processo judicial. Nesse sentido, a atuação da perícia digital revela-se essencial para a verificação técnica da autenticidade de vídeos, áudios e imagens, funcionando como instrumento de garantia da regularidade processual e da verdade material.

Observou-se, ainda, que, apesar do potencial benéfico dos deepfakes, os riscos associados à sua utilização indevida exigem vigilância contínua e resposta institucional adequada. A crescente sofisticação dessa tecnologia dificulta sua identificação, o que reforça a necessidade de desenvolvimento de ferramentas específicas e de capacitação de profissionais especializados na área.

Dentre as medidas recomendadas, destacam-se: a criação de um marco legal específico para regular e punir o uso ilícito de deepfakes; a obrigatoriedade de apresentação de laudos técnicos que atestem a autenticidade das provas digitais; e a implementação de centros especializados em perícia forense digital. Soma-se a isso a necessidade de políticas públicas voltadas à educação digital, com foco na formação de operadores do Direito e da sociedade civil sobre os riscos e mecanismos de prevenção associados ao uso dessa tecnologia.

A partir da integração entre inovação tecnológica, aprimoramento legislativo e qualificação dos profissionais do sistema de justiça, torna-se possível enfrentar os desafios impostos pelos deepfakes, assegurando a efetividade dos direitos fundamentais e a legitimidade das decisões judiciais em um cenário cada vez mais digitalizado.

Referências

SILVA, Germano Marques, Curso de Processo Penal, vol. 1, 6.ª ed., Verbo, 2010, p. 64. SILVA, Germano Marques, Curso de Processo Penal, vol. 1, 6.ª ed., Verbo, 2010, p. 105. SILVA, Germano Marques, Direito Processual Penal Português, Do Procedimento (Marcha do Processo), vol. 3, Universidade Católica Editora, 2015, p. 212. Acessado: 1 de março 2025.

BARFIELD, Woodrow e Ugo Pagallo, Advanced Introduction to Law and Artificial Inte lligence, Elgar, 2020, p. 51. Acessado: 1 de março 2025

KIETZMANN, Jan, Linda W. Lee, Ian P. McCarthy, Tim C. Kietzmann, “Deepfakes: Trick or treat?”, in Business Horizons, volume 63, número 2, 2020, p. 145. Acessado: 1 de março 2025.

WHEELER, Thomas H., Phototruth or Photofiction?: Ethics and Media Imagery in the Digital Age, Routledge, 2002, p. 15. Acessado: 1 de março 2025

KIETZMANN, Jan, Linda W. Lee, Ian P. McCarthy, Tim C. Kietzmann, “Deepfakes: Trick or treat?”, in Business Horizons, volume 63, número 2, 2020, pp. 136-137. Acessado: 3 de março 2025.

SLOOT, Bart van der, Yvette Wagensveld, “Deepfakes: regulatory challenges for the synthetic society”, in Computer Law & Security Review, volume 46, setembro 2022, p. 1. 10. GUI, Jie, Zhenan Sun, Yonggang Wen, Dacheng Tao, Jieping Ye, “A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications”, in arXiv, 2020, p. 2. 11. Acessado: 5 de março 2025

GOODFELLOW, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David War de-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, “Generative Adversarial Networks”, in arXiv, 2014, p. 140. Acessado: 5 de março 2025

HONG, Yongjun, Uiwon Hwang, Jaeyoon Yoo, Sungroh Yoon, “How Generative Adversarial Networks and Their Variants Work: An Overview”, in arXiv, 2019, pp. 1-2. 13. GUI, Jie, Zhenan Sun, Yonggang Wen, Dacheng Tao, Jieping Ye, “A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications”, in arXiv, 2020, p. 1. 14. Acessado: 6 de março 2025.

GOODFELLOW, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David War de-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, “Generative Adversarial Networks”, in arXiv, 2014, p. 139. 15. Acessado: 15 de março 2025.

GOODFELLOW, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David War de-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, “Generative Adversarial Networks”, in arXiv, 2014, p. 139. 16. Acessado: 15 de março 2025.

TREND MICRO RESEARCH, United Nations Interregional Crime and Justice Research Institute (UNICRI), Europol’s European Cybercrime Centre (EC3), Malicious Uses and Abuses of Artificial Intelligence, Trend Micro Research, 2020, pp. 53-54.

Acessado: 20 de março 2025.

TREND MICRO RESEARCH, United Nations Interregional Crime and Justice Research Institute (UNICRI), Europol’s European Cybercrime Centre (EC3), Malicious Uses and Abuses of Artificial Intelligence, Trend Micro Research, 2020, p. 54. 18.

Acessado: 20 de março 2025.

FLORIDI, Luciano, “Artificial Intelligence, Deepfakes and a Future of Ectypes”, in Philosophy & Technology, volume 31, 2018, p. 320. 19. Acessado: 20 de março 2025

TREND MICRO RESEARCH, United Nations Interregional Crime and Justice Research Institute (UNICRI), Europol’s European Cybercrime Centre (EC3), Mali cious Uses and Abuses of Artificial Intelligence, Trend Micro Research, 2020, p. 61. Acessado: 20 de março 2025.

Deepfakes e a Inteligência Artificial: O Papel do Direito Digital no Combate a Fake News no Âmbito Eleitoral, Civil, Penal e Administrativo https://www.jusbrasil.com.br/artigos/deepfakes-e-a-inteligencia-artificial-o-papel-do-

direito-digital-no-combate-a-fake-news-no-ambito-eleitoral-civil-penal-e-administrativo/1194596401. Acessado: 10 de maio2025

Inteligência Atificial no Direito: Impactos e Desafios no Brasil https://legale.com.br/blog/inteligencia-artificial-no-direito-impactos-e-desafios-no-brasil/. Acessado: 20 de maio 2025.

https://inovalegal.org/a-legalidade-das-deepfakes-e-seus-reflexos-na-protecao-de-dados-sensiveis-e-nos-direitos-da-personalidade. Acessado: 20 de maio 2025.