I INTERNATIONAL EXPERIENCE PERUGIA - ITÁLIA

INTELIGÊNCIA ARTIFICIAL: DESAFIOS DA ERA DIGITAL I

ANTÔNIO CARLOS DINIZ MURTA

ANA ELIZABETH LAPA WANDERLEY CAVALCANTI

I61

Inteligência Artificial: Desafios da Era Digital I [Recurso eletrônico on-line] organização CONPEDI

Coordenadores: Ana Elizabeth Lapa Wanderley Cavalcanti, Antônio Carlos Diniz Murta. – Florianópolis: CONPEDI, 2025.

Inclui bibliografia
ISBN: 978-65-5274-095-3
Modo de acesso: www.conpedi.org.br em publicações
Tema: Inteligência Artificial e Sustentabilidade na Era Transnacional

1. Direito – Estudo e ensino (Pós-graduação) – Encontros Internacionais. 2. Inteligência Artificial. 3. Desafios da Era Digital. I International Experience Perugia – Itália. (1: 2025 : Perugia, Itália).

CDU: 34

Conselho Nacional de Pesquisa
e Pós-Graduação em Direito Florianópolis
Santa Catarina – Brasil
www.conpedi.org.br

I INTERNATIONAL EXPERIENCE PERUGIA - ITÁLIA

INTELIGÊNCIA ARTIFICIAL: DESAFIOS DA ERA DIGITAL I

Apresentação

APRESENTAÇÃO DOS ARTIGOS

O Grupo de Trabalho INTELIGÊNCIA ARTIFICIAL: DESAFIOS DA ERA DIGITAL I teve seus trabalhos apresentados nas tardes dos dias 29 e 30 de maio de 2025, durante I INTERNATIONAL EXPERIENCE PERUGIA - ITÁLIA, realizado na cidade de Perugia – Itália, com o tema INTELIGÊNCIA ARTIFICIAL E SUSTENTABILIDADE NA ERA TRANSNACIONAL. Os trabalhos abaixo elencados compuseram o rol das apresentações.

INTELIGÊNCIA ARTIFICIAL: UM NOVO PARADIGMA PARA O PODER JUDICIÁRIO E A REVOLUÇÃO DA JUSTIÇA CONTEMPORÂNEA E DO FUTURO de Eunides Mendes Vieira: Este artigo propõe uma reflexão crítica sobre os impactos da IA no funcionamento da Justiça. Defende que a tecnologia pode reduzir a morosidade e aumentar a previsibilidade das decisões, mas alerta para riscos como viés algorítmico e perda da imparcialidade. Fundamentado em revisão bibliográfica, o texto propõe diretrizes éticas para a adoção da IA no Judiciário, com foco na manutenção dos direitos fundamentais e da equidade no tratamento processual.

A INTELIGÊNCIA ARTIFICIAL NOS TRIBUNAIS: REGULAÇÃO, DESAFIOS E ACCOUNTABILITY de Lais Gomes Bergstein, Douglas da Silva Garcia, Ingrid Kich Severo: O artigo analisa o impacto da inteligência artificial (IA) no Poder Judiciário, destacando sua introdução como mecanismo de automação e celeridade processual. Explora o programa Justiça 4.0 do CNJ, a Plataforma Digital do Poder Judiciário Brasileiro e os marcos regulatórios, como as Resoluções CNJ nº 332 e 335/2020. O texto problematiza a necessidade de governança, transparência e segurança jurídica, especialmente diante da terceirização tecnológica e do uso de dados em nuvem. Conclui-se que o uso da IA deve estar atrelado à ética e à accountability, com observância aos direitos fundamentais.

O USO DA INTELIGÊNCIA ARTIFICIAL NO DIREITO: HARD CASES de Maria de Fátima Dias Santana, Hélcia Macedo de Carvalho Diniz e Silva: O estudo analisa o uso da IA na resolução de hard cases à luz da teoria do Juiz Hércules de Ronald Dworkin. Argumenta que a IA pode contribuir para a celeridade e racionalidade das decisões, mas não substitui a

capacidade de ponderação e interpretação do julgador humano. Traz como exemplo o Projeto VICTOR do STF e propõe que a IA seja usada como instrumento auxiliar, preservando a dimensão humanística da Justiça.

INTELIGÊNCIA ARTIFICIAL E A TRADUÇÃO E GERAÇÃO DE TEXTOS JURÍDICOS de Vanessa Nunes Kaut, Bruno Vinícius Stoppa Carvalho: O texto discute a aplicação de modelos de linguagem (LLMs), como o ChatGPT, na geração e tradução de textos jurídicos. Ressalta o potencial de democratização da escrita jurídica, mas alerta para os riscos à confidencialidade, à autenticidade e à qualidade argumentativa. Aponta que, embora esses sistemas aumentem a produtividade, sua utilização exige regulação adequada, com limites éticos e respeito ao dever de sigilo profissional. O artigo sustenta a importância da supervisão humana e da criação de marcos regulatórios compatíveis com os princípios do Direito.

INTELIGÊNCIA ARTIFICIAL, FISCALIZAÇÃO E CONFORMIDADE TRIBUTÁRIA: DESAFIOS PARA A JUSTIÇA FISCAL de Alexandre Naoki Nishioka, Giulia Ramos Dalmazo: O texto investiga a aplicação da IA na detecção de fraudes fiscais e na conformidade tributária, evidenciando um paradoxo: o mesmo instrumento que fortalece o Fisco também é usado para planejamento tributário abusivo. Analisa a adoção de ferramentas como o SISAM e os desafios éticos e distributivos da automação fiscal. Conclui que é necessário criar estruturas de regulação que conciliem eficiência arrecadatória com justiça fiscal e responsabilidade social.

LIMITES DO CONSENTIMENTO PARENTAL NA PROTEÇÃO DA PRIVACIDADE DOS DADOS PESSOAIS DAS CRIANÇAS NA INTERNET de Gisele Gutierrez De Oliveira Albuquerque: Analisa os desafios jurídicos do consentimento parental no uso de dados de crianças em ambiente digital. Argumenta que a atuação dos pais deve respeitar o princípio do melhor interesse da criança e que o Estado pode e deve impor limites protetivos. Examina normas internacionais e nacionais e conclui pela necessidade de harmonização entre autonomia parental, inovação tecnológica e proteção da infância, principalmente no que tange à coleta e uso de dados pelas plataformas digitais.

PROTEÇÃO DE DADOS PESSOAIS DE CRIANÇAS E ADOLESCENTES E A INTELIGÊNCIA ARTIFICIAL: UM OLHAR SOB A PERSPECTIVA DA LEGISLAÇÃO BRASILEIRA de Ana Elizabeth Lapa Wanderley Cavalcanti, Patrícia Cristina Vasques De Souza Gorisch: Este artigo trata dos desafios específicos enfrentados na proteção de dados pessoais de crianças e adolescentes no contexto da IA e das redes digitais. Analisa a legislação brasileira, como a LGPD, o ECA e a Constituição Federal, destacando a centralidade do princípio do melhor interesse da criança. Argumenta que é necessário rever o

papel do consentimento parental frente à hipervulnerabilidade infantojuvenil e propõe medidas de educação digital, regulação e fiscalização mais efetivas, com foco na proteção integral desse grupo.

QUEM OLHA PELOS SEUS OLHOS? UMA ANÁLISE SOBRE A PROTEÇÃO DE DADOS E A PROVA DE PERSONALIDADE de Edith Maria Barbosa Ramos, Pedro Gonçalo Tavares Trovão do Rosário, Pastora Do Socorro Teixeira Leal: Explora a relação entre a proteção de dados pessoais e a noção de personalidade jurídica, especialmente no contexto da vigilância digital e do uso de IA. Retoma o debate sobre o direito à privacidade a partir de sua construção histórica e reforça que a proteção dos dados é expressão direta da dignidade da pessoa humana. A obra destaca o conceito de "prova de personalidade" como um novo paradigma jurídico, que busca assegurar o controle individual sobre as informações pessoais em tempos de capitalismo de dados.

PRECISAMOS FALAR SOBRE A DISCRIMINAÇÃO ALGORÍTMICA NAS RELAÇÕES DE CONSUMO de Dennis Verbicaro Soares, Loiane da Ponte Souza Prado Verbicaro: O texto aborda como algoritmos utilizados em plataformas digitais e ferramentas de IA têm reproduzido e intensificado práticas discriminatórias contra grupos vulneráveis. Explica que a predição comportamental, quando não supervisionada, pode resultar em decisões automatizadas excludentes, violando o princípio da isonomia. Propõe a criação de um Direito Antidiscriminatório aplicado à tecnologia, bem como a implementação de políticas públicas e marcos regulatórios que evitem a colonização algorítmica do consumidor e assegurem o respeito à dignidade nas relações de consumo.

PERSPECTIVAS E DESAFIOS À GOVERNANÇA TRANSNACIONAL DA INTERNET NA SOCIEDADE DIGITAL de Vanessa De Ramos Keller: O artigo propõe uma reflexão crítica sobre a ausência de uma governança global eficaz da internet. Defende que, em um mundo interconectado, não há mais espaço para ações unilaterais, sendo necessária a criação de um sistema de governança transnacional. Ressalta-se o papel das big techs e a necessidade de coordenação internacional para garantir direitos digitais, proteção de dados, liberdade de expressão e combate à desinformação. A obra argumenta que a sociedade digital demanda novos paradigmas jurídicos e políticos capazes de enfrentar os desafios da era informacional.

OS LIMITES BIOLÓGICOS E COGNITIVOS DA INTELIGÊNCIA ARTIFICIAL: UMA ANÁLISE SOBRE A SUSTENTABILIDADE INERENTE AOS IMPACTOS DA IA NA CAPACIDADE SÓCIO-COGNITIVA HUMANA de Aulus Eduardo Teixeira de Souza: Com abordagem interdisciplinar, o artigo discute as barreiras físicas, cognitivas e éticas que limitam a capacidade da inteligência artificial em simular a cognição humana. Contrapõe a

eficiência energética e adaptabilidade do cérebro humano com os altos custos computacionais e a rigidez dos sistemas de IA. Ressalta que a ausência de consciência subjetiva e de empatia torna a IA inadequada para decisões sensíveis. Conclui pela importância de reconhecer os limites biológicos da IA como base para um desenvolvimento tecnológico mais sustentável e responsável.

ORGANIZAÇÕES CRIMINOSAS: A IMPORTÂNCIA DA INTELIGÊNCIA ARTIFICIAL NO ENFRENTAMENTO DO CRIME ORGANIZADO deRoberta Priscila de Araújo Lima, Alice Arlinda Santos Sobral, Raylene Rodrigues De Sena: O estudo destaca o papel da inteligência artificial como aliada estratégica no combate ao crime organizado. Após um panorama da evolução normativa brasileira sobre o tema, especialmente com a Lei 12.850 /2013, o texto evidencia como a IA pode ser utilizada em ações policiais e de inteligência, facilitando a análise de grandes volumes de dados, identificando padrões e prevenindo crimes. A pesquisa conclui que o uso responsável e regulamentado da IA pode fortalecer a segurança pública e otimizar as ações de combate ao crime organizado, respeitando garantias legais e direitos fundamentais.

NEURODIREITOS E INTELIGÊNCIA ARTIFICIAL: MAPEAMENTO PROTETIVO DOS DIREITOS HUMANOS E FUNDAMENTAIS NA SOCIEDADE 4.0 de Simone Gomes Leal, Olivia Oliveira Guimarães: Explora o conceito de neurodireitos como nova categoria de direitos humanos frente à interface entre IA e neurotecnologia. Destaca os riscos à dignidade humana, à identidade e à privacidade mental causados por tecnologias que acessam ou modulam o cérebro. Enfatiza o papel do constitucionalismo digital na proteção desses direitos, propondo sua positivação nas legislações nacionais e internacionais como forma de preservar a integridade do sujeito frente à máquina.

VIESES ALGORÍTMICOS E RECONHECIMENTO FACIAL de Pedro Henrique do Prado Haram Colucci, Sergio Nojiri: Analisa o caso do Projeto Vídeo-Polícia Expansão, implantado na Bahia, e seus efeitos discriminatórios. O artigo mostra como sistemas de reconhecimento facial produzem falsos positivos, especialmente contra pessoas negras, e denuncia a ausência de regulamentação e de auditorias obrigatórias. Propõe modelos internacionais para nortear a regulação brasileira.

IA NA GESTÃO MIGRATÓRIA: INCLUSÃO DIGITAL OU FERRAMENTA DE EXCLUSÃO? de Patricia Cristina Vasques De Souza Gorisch, Ana Elizabeth Lapa Wanderley Cavalcanti: Examina a crescente utilização da IA em políticas migratórias, como triagem de pedidos de refúgio, monitoramento de fronteiras e identificação de migrantes. Denuncia que, embora a tecnologia possa facilitar o acesso a serviços, também é usada para

vigilância e exclusão de grupos vulneráveis. O texto propõe uma regulação ética e baseada nos direitos humanos.

A CIDADANIA ELETRÔNICA DO HOMO DIGITALIS: PERSPECTIVAS JURÍDICAS À LUZ DO REGULAMENTO EU 2024/1689 SOBRE INTELIGÊNCIA ARTIFICIAL de Olivia Oliveira Guimarães, Helen Caroline Cardoso Santos, Lucas Gonçalves da Silva: Trabalha a Inteligência Artificial sob o aspecto da regulação europeia, tendo como base a questão da cidadania digital.

DECISÕES AUTOMATIZADAS E COGNIÇÃO HUMANA: O PAPEL DA INTELIGÊNCIA ARTIFICIAL NO PROCESSO DECISÓRIO JUDICIAL de Sergio Nojiri, Luiz Guilherme da Silva Rangel: Tratando de questões atinentes ao uso da Inteligência Artificial em decisões judiciais.

TRANSAÇÃO NA REFORMA TRIBUTÁRIA COMO MEDIDA DE DESJUDICIALIZAÇÃO de Tammara Drumond Mendes, Antônio Carlos Diniz Murta, Renata Apolinário de Castro Lima.

VEDAÇÃO AO CONFISCO DA PROPRIEDADE ÚNICA QUE ATENDE A FUNÇÃO SOCIAL de Tammara Drumond Mendes, Antônio Carlos Diniz Murta, Renata Apolinário de Castro Lima.

Após duas tardes de intensos debates sobre os temas apresentados, foram encerrados os trabalhos do GT com a elaboração de uma síntese que se chamou de Carta de Perúgia.

Os temas demonstram a abrangência e amplitude do tema que é de grande interesse da ciência jurídica e que permite uma profícua produção acadêmica nacional e internacional. Importante lembrar que os pesquisadores presentes no GT estão vinculados aos mais diversos programas de pós-graduação em Direito, demonstrando a importância de debates como os ocorridos nos dias 28, 29 e 30 de maio de 2025, na cidade de Perúgia – Itália.

Nota-se preocupação de todos quanto à regulação da Inteligência artificial, mormente para que não só, numa visão meramente apocalíptica, se torne um instrumento de maior concentração de poder nas mãos de grandes grupos - big techs - e manipulação comportamental, mas também não possa ser a médio prazo um elemento que possa reduzir a liberdade e autonomia humana no pensar e evoluir seja em questões técnicas seja em questões sociais/filosóficas. Não existem dúvidas que enfrentamos uma nova realidade sem embargo de ser virtual e não materializada que vai exigir da comunidade internacional ou de

cada um de nós adequação para um fenômeno que não pode ser impedido; mas pode ser, a partir de um maior aprofundamento sobre seu poder e efeitos na sociedade, melhor assimilado sem que percamos, sendo otimista, o que nos torna humanos.

Diante da diversidade de temas e das pesquisas de grande qualidade apresentadas neste evento, recomendamos que operadores do direito em todas as suas funções leiam os trabalhos aqui apresentados.

Coordenadores:

Antônio Carlos Diniz Murta

Universidade FUMEC

acmurta@fumec.br

Ana Elizabeth Lapa Wanderley Cavalcanti

Universidade Presbiteriana Mackenzie

ana.cavalcanti@mackenzie.br

UM MAPEAMENTO DE ESTRATÉGIAS E POLÍTICAS PÚBLICAS ADOTADAS NO COMBATE AOS IMPACTOS DO DEEPFAKE ENFRENTADOS NO PROCESSO GARANTIDOR DE INTEGRIDADE DA INFORMAÇÃO: REGULAÇÕES, DEMOCRACIA E DESINFORMAÇÃO.

A MAPPING OF STRATEGIES AND PUBLIC POLICIES ADOPTED TO COMBAT THE IMPACTS OF DEEPFAKE ON THE PROCESS OF ENSURING INFORMATION INTEGRITY: REGULATIONS, DEMOCRACY, AND DISINFORMATIO

Yuri Nathan da Costa Lannes
Lais Faleiros Furuya
Laís Reis Araújo Nazaré

Resumo

O presente artigo científico discute as possíveis soluções mitigadoras do deepfake em face do impactos causados à integridade informacional. O estudo aborda os temas relacionados à verificação da veracidade dos conteúdos disseminados massivamente na internet, como fake news, infodemia, pós-verdade e a crise da democracia em razão da falsa liberdade de expressão dos usuários da internet. Com uma abordagem metodológica qualitativa e quantitativa, o estudo é realizado com revisão em artigos cinetíficos que tratam tecnicamente sobre o deepfake e a integridade da informação, além da análise de relatórios que abordam sobre os casos envolvendo a manipulação de imagens, vídeos e audios. Faz parte da metodologia as investigações de boletins e informativos que contenham quantitativos de usuários da internet que possuem habilidades em verificar a veracidade das informações da internet; números daqueles que compartilham informações sem identificar o referido conteúdo e ainda a quantidade dos que dispõem de conhecimento sobre a existência de deepfake presentes na internet. O objetivo é mapear políticas publicas e iniciativas privadas capazes de detectar conteúdos manipulados com tecnologia de inteligência artificial e garatir que as informações sejam compartilhadas na sua forma integra. Justifica-se o pesente estudo em face da crise na democracia causado pela falsa liberdade de expressão e a polarização na internet, das quais são intensificados pelo uso de deepfake como extensão da fake news.

Palavras-chave: Integridade informacional, Deepfake, Desinformação, Políticas públicas, Fake news

Abstract/Resumen/Résumé

This scientific article discusses possible mitigating solutions for deepfake in light of its impacts on informational integrity. The study addresses topics related to verifying the authenticity of massively disseminated online content, such as fake news, infodemia, post-truth, and the crisis of democracy due to the false freedom of expression granted to internet users. Using a qualitative and quantitative methodological approach, the research is conducted through a review of scientific articles that technically address deepfake and

information integrity, as well as an analysis of reports that discuss cases involving the manipulation of images, videos, and audio. The methodology also includes investigations of bulletins and reports containing data on internet users' ability to verify the authenticity of online information, the number of individuals who share content without verifying its accuracy, and the percentage of users aware of the existence of deepfake technology. The objective is to map public policies and private initiatives capable of detecting AI-manipulated content and ensuring that information is shared in its original, unaltered form. This study is justified by the crisis in democracy caused by false freedom of expression and online polarization, both of which are intensified by the use of deepfake as an extension of fake news.

Keywords/Palabras-claves/Mots-clés: Informational integrity, Deepfake, Disinformation, Public policies, Fake news

1 INTRODUÇÃO

O presente artigo científico terá como discussão principal a análise de desafios enfrentados no combate à integridade informacional em face do emprego das técnicas de *deepfake* como extensão da crescente realidade de fake news disseminadas na internet. Esta temática tem se tornado relevante em face da progressividade de casos envolvendo a divulgação de imagens, vídeos e áudios manipulados com inteligência artificial.

Esta pesquisa justifica-se pelos impactos da *deepfake* na era da intertnet, e a necessidade de estratégias e soluções que devem ser adotadas em razão da quantidade massiva de conteúdos manipulados pela tecnologia. A falta de regulamentação específica sobre o tema motiva o presente estudo a busca por planos que se materializam como solução em face das ameaças à integridade informacional e a intensificação da desinformação, ambos decorrentes de fake news intensificadas pelo *deepfake*. Para tanto o estudo se delimita a explorar o que é a integridade da informação e como as *deepfakes* e a infodemia estão gerando uma ciberpolarização, sendo este cenário influenciado significativamente pelas tecnologias de informação e comunicação. A análise se estende nas regulamentações que fomentam estratégias opositoras ao compartilhamento de informações manipuladas tecnologicamente.

Nota-se que o problema central deriva da combinação da internet com as tecnologias disruptivas que estão tornando o acesso à informação cada vez mais intenso. Apesar desses meios serem garantidores de uma liberdade de expressão, a transmissão de conteúdos diversos está se tornando cada vez mais rápida, podendo eles serem verdadeiros ou não. O problema é que na realidade da internet há a presença de usuários que absorvem uma quantidade massiva de informações das quais são palatáveis aos seus valores religiosos e de senso comum.

Por consequência esses conteúdos são filtrados sem que haja a devida verificação da sua integridade, com a possibilidade do seu compartilhamento e a criação de um ambiente favorável para a desinformação. Este contexto se agrava quando as tecnologias de *deepfake* se inserem, alterando uma informação e tornando-a impossível de distinguir com aquelas que são verdadeiras. Em face desse cenário, a pergunta que se pretende responder nesta pesquisa é: quais os desafios que a integridade informacional pode enfrentar com o uso do *deepfake*? E quais estratégias e soluções podem ser visualizadas para combater essa habilidade tecnológica provocadora de fake news e desinformação?

Deste modo, em termos gerais tem-se como objetivo do estudo analisar formas e iniciativas que podem ser empregadas pela sociedade, terceiro setor, e entes públicos e privados

no combate das técnicas de *deepfake* que influenciam diretamente na garantia da integridade da informação. Na perspectiva dos objetivos específicos, enumera-se o exame da integridade da informação e quais abordagens estão sendo usadas pelos países, incluindo o Brasil diante desta nova agenda de estudo; como as novas tecnologias de inteligência artificial manipuladoras de conteúdos como o *deepfake* inserem-se como barreira da veracidade das informações e, apuração de planos públicos e privados que surgem como solução no combate ao compartilhamento de conteúdo alterados por inovações tecnológicas.

Para realizar o presente estudo, será realizada uma abordagem qualitativa e quantitativa, considerando o uso de revisões bibliográficas, investigação de casos concretos e análise de relatórios empíricos sobre a realidade das fake news e *deepfake* presentes na internet. Deste modo, cartas de recomendação, informativos, documentos técnicos, legislações, artigos científicos serão usados como fonte primária, enquanto matérias de jornais entrará como fontes secundárias, tendo em vista a abordagem de relatos reais envolvendo o tema em estudo.

Por fim, o artigo será estruturado em três capítulos, abordados da seguinte forma: destrinchamento do tema integridade da informação, verificação de termos correlatos, como pós-verdade, infodemia, desinformação e fake news e, mapeamento de posições políticas quanto à nova agenda de pesquisa, posteriormente a compreensão sobre do termo *deepfake*, seu funcionamento, os casos relevantes envolvendo esta habilidade e análise projetos de lei envolvendo o referido termo e, por fim, um estudo do caso do Governador do Estado de São Paulo Tarcisio de Freitas, vítima de *deepfake* com o mapeamento de possíveis soluções ao combate desta técnica na Era da Informação.

2.1 INTEGRIDADE INFORMACIONAL E OS EFEITOS DA DESINFORMAÇÃO

As tecnologias de informação e comunicação (TIC) tornaram o acesso à internet mais fácil com dados e informações sendo transmitidas em massa e a todo momento. Porém neste cenário há uma preocupação se o conjunto informacional é propagado na sua forma real ou modificada, razão pela qual se faz necessário analisar neste capítulo a importância dada à necessidade de garantir a integridade da informação e os efeitos quando a veracidade do conteúdo não é resguardada.

Pesquisa realizada pelo Senado e a Câmara dos Deputados (2019) identificou que a cada duas mil pessoas, mil e quinhentas usam o aplicativo WhatsApp como principal fonte de informação. Nota-se que este número equivale a 79% das pessoas participantes do estudo. Com essa informação, deve-se levantar um primeiro ponto, que é a falsa sensação de liberdade pela

quantidade de informações disponíveis. A problemática dessa questão é que as inovações colocam à disposição uma quantidade massiva de dados, dos quais são incapazes de serem filtrados e averiguados intactamente pelo receptor (Araujo, Santos, 2024, p. 13). A confirmação está numa pesquisa da Kaspersky em 2021, identificando que 3 em cada 4 pessoas da América Latina se sente sobrecarregada de informações expostas na internet (Kaspersky, 2021))

O resultado é que essa impossibilidade humana de verificar a natureza de todos os conteúdos absorvidos gera um consumo excessivo de informações das quais nem sempre são verdadeiras, traduzindo à uma liberdade fictícia de acesso amplo a todos os tipos de conteúdo. Ramonet (*apud* Araujo, Santos, 2024, p. 13) caracteriza este cenário como "censura democrática". Neste caso, não há barreiras e impedimentos na liberdade de ir e vir, mas sim um volume de informações que são absorvidas pelos receptores sem a devida integridade sobre sua matéria.

Uma segunda análise a ser realizada sobre essa falsa liberdade refere-se ao manuseio das informações pelas novas tecnologias disruptivas. Neste sentido, como ocorre com as redes socais, o emprego de ferramentas com inteligência artificial e operação algorítmica trabalham com os dados norteando quais e quantas informações vão estar dispostas ao usuário. Apesar de acreditar ter o controle e o poder de direcionar os dados que deseja absorver, o indivíduo é submetido pelas tecnologias que direcionam as informações, sendo elas íntegras ou não. Por consequência nasce a nova realidade da crise da democracia e o processo de formação do cidadão brasileiro em face da confiança excessiva sobre as notícias que lhe são compartilhadas (Araujo, Santos, 2024, p. 16)

Em terceiro lugar há outro ponto problemático que é resultado destes dois primeiros: o enraizamento da nova "cultura da pós-verdade" (Araújo, 2024, p. 2017). A aderência a este novo termo indica a potencialidade dos indivíduos estarem formando suas opiniões fundamentadas em crenças religiosas e sentimentos subjetivos próprios. Isso caracteriza como um desafio no processo de formação das convicções, pois não se dá a devida importância em analisar se os dados são verídicos ou não.

O autor Carlos Araújo (2024, p. 217) caracteriza essa nova cultura com três pilares. O primeiro é a disseminação de informações com conteúdo modificado a partir do auxílio das novas tecnologias disruptivas. A comprovação está no estudo realizado pelo Instituto Locomotiva. A pesquisa foi realizada entre os dias 15 e 20 de fevereiro de 2024, com 1.032 entrevistas à indivíduos da faixa etária maior de 18 anos. O resultado foi que 90% dos entrevistados assumiram que acreditaram em conteúdos falsos e 64% acreditam que estes conteúdos derivam de tecnologias de inteligência artificial (Instituto Locomotiva, 2024).

O segundo pilar é a verificação da integridade da informação e a possibilidade de identificar a veracidade da notícia recebida (Araújo, 2024, p. 217). O Centro Regional de Estudo para Desenvolvimento da Sociedade e da Informação identificou no ano de 2024 o número de usuários de internet por tipo de habilidade digital. Dentre as habilidades analisadas, convém indicar que uma delas é a destreza em verificar a veracidade de uma informação encontrada na internet. No cenário de comparações, o que chama atenção é que dentre os analfabetos usuários de internet no Brasil, apenas 5% realiza a referida verificação. Paralelamente, depara-se que 80% dos usuários de internet no Brasil que tem um ensino superior tem a habilidade mencionada (Cetic.br, A-1, 2024).

Para confrontar estes dados, é interessante pontuar que conforme a Pesquisa Nacional por Amostra de Domicílios Contínua (PENAD) realizada pelo Instituto Brasileiro de Geografia e Estatística, há apenas 5,6% de indivíduos com 15 anos ou mais que são analfabetos. Por outro lado, indicou-se a taxa de 15,4% de pessoas com 60 anos ou mais que se enquadram como analfabetos (IBGE, 2023). Portanto, mostra-se a existência de um grupo de vulneráveis quanto ao tema da integridade informacional e desinformação.

Por fim, o terceiro pilar é o número de pessoas que não realizam a conferência das informações, procedendo com a divulgação delas sem realizar a devida checagem (Araújo, 2024, p. 217). Com um olhar nos números, em 2022, o *Poynter Institute* juntamente com a Google identificou que de dez pessoas, quatro recebem conteúdo alterado todos os dias. Nesta mesma pesquisa, dos 8,5 mil entrevistados, pelo menos 43% sinalizaram que já passaram notícias das quais descobriram seu conteúdo falso após o envio (Guimarães, Rodrigues, 2022). Para complementar, o Comitê Gestor da Internet no Brasil juntamente com a Unesco realizou um estudo com o grupo de indivíduos de 9 a 17 anos que usam a internet. De 2.607 entrevistados, no período de junho e outubro de 2022, constatou que 43% deste grupo analisado não tem habilidade de verificar a veracidade da informação exposta na internet (CGI.br, 2023).

A partir destas três condições, é possível identificar a falta de preocupação se a notícia é falsa ou não, mas apenas apoiar em verdades fundadas em emoções, informações agradáveis de absorver e pré-conceitos religiosos. Em face do poder desta nova cultura, é notável a preocupação do Brasil e dos países do G20 sobre a desinformação e a integridade da informação, sendo este último o novo termo adotado nas recentes discussões.

Em relação à desinformação, ela é um efeito da cultura pós-verdade e da falsa sensação de liberdade em meio à quantidade massiva de informação disponível na internet. A partir do sentido amplo deste termo, derivam-se três subdivisões. A primeira é a “desinformação” (Araújo, 2024, p. 218), que é aquela notícia falsa repassada com vontade de gerar prejuízos.

Em relação à segunda, é a "informação falsa ou errônea" (2024, p. 218), que é a notícia falsa divulgada sem a vontade de prejudicar outrem. E por fim, a "informação maliciosa" (2024, p. 2018) que é uma informação verdadeira, mas divulgada de forma deturpada para gerar prejuízos.

Dessas derivações, há um foco em comum que é o termo *fake News,* conceituado por "toda informação difundida por meios de comunicação que se disfarçam de veículos jornalísticos e que difundem informação comprovadamente incorreta para enganar seu público" (Alcott, Gentzkov *apud* Santos, Araújo, 2024, p. 15). A diferença com a desinformação no seu conceito amplo é que aquele termo se trata de um conteúdo falso, enquanto este refere-se a um espaço com barreiras na circulação livre de informações verídicas (Tribunal Regional Eleitoral, 2023).

Somado a este conceito atrelado à desinformação, o termo infodemia também se correlaciona aos desafios da integridade da informação. Nesse sentido, a palavra define-se como o compartilhamento massivo de informações falsas em uma grande área geográfica. A consequência são indivíduos que recebem mais notícias desvirtuadas do que verdadeiras. Por lógica, o efeito é prejudicial, caracterizando na desconfiança sobre os demais conteúdos na internet que são polêmicos em sua essência, gerando a ciberpolarização (Araújo, 2024, p. 2019).

Por outro lado, o termo integridade da informação está cada vez mais aparecendo em agendas de discussões realizadas por entes públicos e pesquisadores da área. Para este estudo, considera que o conceito deste termo é "como a confiabilidade, o equilíbrio e a completude das informações às quais os cidadãos têm acesso, relacionadas aos termos políticos" (Araújo, 2024, p. 2013).

No Brasil, há uma preocupação significativa sobre essa integridade. A confirmação está no evento realizado no mês de abril de 2024, denominado de Conferência Livre: Ciência no Combate à Desinformação, organizado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) em parceria com outras instituições de pesquisa. Neste momento, pesquisadores, docentes e gestores da administração pública discutiram sobre o papel da ciência e da pesquisa nas barreiras da integridade informacional e os seus efeitos negativos, como o caso da desinformação (IBICT, 2024).

Outro marco temporal foi o encontro de investigadores sobre o tema no evento realizado nos meses de abril e maio de 2024, pelo Grupo de Trabalho de Economia Digital do G20. Trata-se de um grupo de pesquisadores, cujo centro de estudos direciona-se a quatro eixos diferentes, sendo eles: integridade da informação, conectividade significativa, governo digital e inteligência artificial. Sobre o evento, o tema central das discussões foi a "Integridade da

Informação e Confiança no Ambiente Digital". A ideia do evento é que a desinformação e a veracidade das informações da internet fosse palco de discussões e estudos em busca de soluções garantidoras da referida probidade dos conteúdos compartilhados na internet (Secretaria de Comunicação Social, 2024).

Ainda dentro do ano de 2024, em dezembro, o Brasil associou-se à Recomendação sobre Integridade da Informação elaborado pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE). Em linhas gerais, o documento atribui sugestões direcionadas aos três principais eixos. Todos direcionados a formação de habilidades em indivíduos capazes de desviar da desinformação, a necessidade de a procedência da informação ser transparente e a importância de entes públicos buscar soluções para fortalecer seus sistemas fiscalizadores de veracidade nas informações da internet. (Secretaria de Comunicação Social, 2024).

Além dessas movimentações nacionais sobre a preocupação da integridade da informação em meio às novas tecnologias da informação e comunicação, há organizações que se mobilizam com a mesma preocupação. Um exemplo é o The World Leadership Alliance - Club de Madrid, que em 2018 elaborou o *"Protecting Information Integrity: National and International Policy Options"*. Em tese, o documento classifica a informação como um Direito base para a democracia, considerando que para exercer direitos políticos é necessário o acesso a dados sobre práticas e políticas públicas estatais. Nesse sentido o documento estabelece uma linha tênue entre as ferramentas tecnológicas que surgem como fortalecedoras do Direito à informação, mas que em contrapartida são ameaças à integridade do conteúdo divulgado. Apesar de facilitar o acesso às notícias, auxiliando na formação de opiniões e facilitando a liberdade de expressão, o meio operacional da disseminação de informações, os interesses comerciais e capitais, os algorítmicos e os seus vieses são deturpadores dos respetivos conteúdos (Araújo, 2024, p. 213).

Um segundo documento que demonstra a mobilização internacional é denominado por "Integridade da informação: construindo o caminho para a verdade, resiliência e a confiança". O conteúdo foi elaborado pelo Programa de Desenvolvimento das Nações Unidas e publicado no ano de 2022. A ideia repassada é que o processo de garantir a integridade da informação é um meio facilitador e impulsionador na tomada de várias decisões. Dentre elas é possível exemplificar mais uma vez com o cenário eleitoral e a supervisão do povo sobre as ações e políticas públicas de um determinado governo. Nesse sentido, as tecnologias de informação e comunicação aparecem mais uma vez, como fomentadoras da disseminação dessas informações. O problema é que este estímulo pode ser positivo, como a solidificação da

democracia, mas negativo, como o uso de inteligências artificiais impulsionadoras de desinformação (Araújo, 2024, p. 214).

Em suma, em ambos os documentos há uma preocupação em comum, que é a presença prejudicial de tecnologias disruptivas somadas à internet. O grande problema é que com a internet e a nova realidade digital, os usuários inseridos na cultura pós-verdade e nos meios de comunicação e veiculação de informação, absorvem com mais facilidade conteúdos agradáveis aos olhos, independente da verificação da verdade. O resultado é a polarização dos indivíduos, em que estes criam suas próprias bolhas. Elas são estimuladas significativamente por notícias falsas criadas, das quais vem sendo aperfeiçoadas pelas tecnologias disruptivas. Esta habilidade de manipular conteúdo a partir do uso de ferramentas de inteligência artificial de forma negativa é que se chama de *deepfake*. Em razão dos impactos causados por este fenômeno no processo garantidor da integridade informacional, nota-se que o próximo capítulo cuidará do seu estudo específico.

2.2 DESAFIOS CAUSADOS PELO DEEP FAKE À INTEGRIDADE DA INFORMAÇÃO

As recentes preocupações com a integridade da informação por parte dos gestores públicos nacionais e internacionais foram detalhadamente demonstradas. São vários conceitos que giram em torno deste tema central, como desinformação, infodemia, pós-verdade, fake News e *deepfake*. Um estudo realizado pelo Instituto de Tecnologia de Massachusetts (MIT) no ano de 2018, identificou que fake news são repassados 70% mais rápido do que informações verídicas (Agência Estado, 2018). A questão é que essas notícias falsas estão sendo aperfeiçoadas cada vez mais por ferramentas tecnológicas. A técnica de manipulação digital dessas informações é o que se denomina *deepfake*, sendo, portanto, este o objeto de discussão do presente capítulo, justificando-se em razão de tal abordagem ser uma ameaça à veracidade do conteúdo disposto na internet quando empregada para gerar prejuízos.

De forma introdutória, o *deepfake* tem como tradução livre a falsificação profunda. Deste modo é possível conceituar como "identidades falsas criadas com deep learning [aprendizagem profunda, por meio de uso maciço de dados] por meio de uma técnica de síntese de imagem humana baseada na inteligência artificial" (Dias *et al*, 2024, p. 7). Sua prática é "combinar e sobrepor imagens e vídeos preexistentes e transformá-los em imagens ou vídeos "originais"" (Dias *et al*, 2024, p. 7). Na prática, o *deep fake* resulta na criação de vídeos, áudios e imagens falsas, retratando cenas que nunca ocorreram de verdade (Kaspersky).

Com um olhar operacional, tecnicamente consiste em uma tecnologia que manipula vídeos, imagens, áudios com uso da metodologia de deep learning e o seu complexo de redes neurais. Com um treinamento constante a partir de um banco de dados formados por rostos e vozes, o conteúdo falsificado torna quase impossível identificar a falta de veracidade. A problemática é que nos navegadores da internet, há um amplo acesso de sistemas computacionais com códigos algorítmicos acessíveis e até mesmo aplicativos facilmente disponíveis para aplicar o *deepfake* (Dias *et al*, 2024, p. 7).

O funcionamento é a partir das chamadas Redes Neurais Generativas Adversárias (GAN's), que operam treinando a partir de modelos que existem para falsificar situações que nunca ocorreram na realidade. Esses modelos de aprendizado profundo são formados por duas redes. A primeira é o "modelo gerador" (Berenguel, Molina, 2022, p. 4), que cria novos conjuntos de dados. Já o segundo é o "modelo discriminador" (Berenguel, Molina, 2022, p. 4), responsável por classificar se o conjunto de novos dados criados coincidem ou não com os dados inseridos para treinamento.

Os pesquisadores Courville *et. al* (*apud* Berenguel, Molina, 2022, p. 4) associam os modelos gerados a falsificadores e os modelos discriminadores aos policiais. Deste modo, ambos os lados estão sempre se debatendo, pois de um lado a polícia precisa identificar se é falso ou não, enquanto do outro lado, os falsificadores estão sempre trabalhando para fazer com que seus produtos falsos sejam indiferenciáveis ao produto de primeira linha. Há processos que são mais simples, que não precisam adequar o formato do rosto por exemplo, mas apenas a velocidade do discurso. Essa técnica mais básica é denominada por cheapfakes, sendo associadas por "falsificações baratas" (Supremo Tribunal Federal, 2024).

Para ter uma dimensão dos efeitos dessa habilidade, a Kaspersky juntamente com a CORPA, realizou uma pesquisa nos meses de fevereiro e março no ano de 2021, entrevistando 2.358 indivíduos da faixa etária de 25 e 65 anos. O resultado apontou que 66% brasileiros negligenciam a possibilidade das suas informações terem sido produzidas ou não com a referida técnica. Além disso, 63% dos indivíduos do Brasil tem um conhecimento mínimo sobre a técnica e 71% deste mesmo grupo não consegue identificar se o vídeo ou imagem teve o emprego dessa habilidade (Kaspersky, 2022).

A questão é que esse termo inserido no contexto de pós-verdade direciona o olhar daquele que está absorvendo a informação, não havendo o interesse em saber se há verdade ou não no conteúdo. No caso das plataformas digitais, os algoritmos presentes nos meios de disseminação de informação identificam quais conteúdos chamam atenção dos seus usuários, tornando um ambiente favorável ao repasse de fake News e de imagens e vídeos manipulados

para tornar o conteúdo mais aderente ao receptor. Para complementar, há ainda o desinteresse e falta de habilidade na verificação de veracidade na informação, o que certamente contribuiu para a formação da ciberpolarização e do ambiente de desinformação. Em suma, a divulgação de fake News associadas com técnicas de *deepfake* são favorecidas pela bolha criada pelo próprio indivíduo ao navegar na internet (Jacoob, 2020, p. 297).

Com uma visão mais ampla é possível afirmar que o uso desta técnica pode ser direcionado a vários tipos de danos, desde produção de provas falsas, coação, vídeos pornográficos até assuntos políticos, como no exemplo anterior. O problema é que quando se trata de questões políticas e ações governamentais, essa habilidade pode ser usada para realizar discursos políticos que nunca ocorreram de verdade. Para agravar, esses conteúdos podem ser repassados com grande facilidade na internet, sem haver qualquer tipo de verificação sobre a sua veracidade. Isso facilita a disseminação de fake News e a integridade da informação que é compartilhada, estimulando cada vez mais a polarização na internet e a noção de pós-verdade (Dias, 2024, p. 9).

Sobre a Fake News, das quais a *deepfake* se deriva, deve-se ressaltar sobre a importância dada a elas em razão das suas ameaças à integridade informacional. A empresa PSafe de cibersegurança realizou um estudo em 2020 levando como referência o número de 131,1 milhões de brasileiros que possuem Android. A pesquisa obteve 70.333 questionários respondidos. O resultado é que 45,94% já receberam fake News sobre assuntos de saúde, 37,80% sobre celebridades e 33,57% sobre política (PSafe, 2020).

Outro resultado que chama atenção é que a plataforma em que mais se encontrou notícias falsas foi o Facebook, com 40.48%. E ainda, 55% já compartilharam uma informação falsa e só depois descobriu a falta de veracidade do seu conteúdo (PSafe, 2020). Ou seja, pode-se dizer que mais da metade realiza o repasse de informações falsas, mas descobre sobre sua natureza logo após. O prejuízo é a influência negativa sobre determinado assunto, resultado possivelmente no fomento à valores negativos, preconceito e intolerância. É essencial considerar o prejuízo à integridade da informação e a solidificação da pós-verdade quando é inserido neste cenário a técnica de *deepfake*s.

Dentro dessa análise, nota-se que a Constituição Federal de 1988 sinaliza em alguns momentos sobre a liberdade de expressão. O artigo 5, inciso IV e IX garante respectivamente a liberdade de manifestação de pensamento e a liberdade de expressão de qualquer que seja a atividade, inclusive de comunicação. Neste mesmo artigo, inciso XIV e no artigo 220 há um resguardo do direito de acesso e exposição da informação em qualquer veículo de comunicação (Brasil, 1988). Com essa base legal, afirma-se que a liberdade é garantidora da democracia,

sendo a tecnologia facilitadora no exercício deste direito e na posição de cidadão. A partir do momento que se insere o cenário de fake News e o uso da técnica em estudo, esta liberdade torna-se ameaçada (Dias, 2024, p. 13).

O relatório *The State Of deepfakes 2024* elaborado pela *Sensity* confirmou a existência de 12 plataformas e sites responsáveis por repassar *deepfakes*. Outro resultado apontado foi que dentre as ferramentas desta técnica, então em maioria aquelas que realizam a troca do rosto por um igual elaborado pela inteligência artificial, com sincronização lábia, reencenação facial e avatares virtuais. Em segundo lugar, estão as técnicas de falsificação de imagens pela IA, enquanto os clones de voz pela mesma técnica de aprendizado de máquina encontram-se em terceiro lugar (Sensity, 2024)

Em casos práticos, pode-se citar o vídeo criado com a imagem de Mark Zuckerberg, CEO da plataforma do Facebook. O conteúdo era basicamente um discurso seu de que a referida *bigtech* supervisiona e coordena o futuro de todos os seus usuários pelo simples controle de dados que são roubados da plataforma e de outras interligadas, como o Instagram. Os dados base para o processamento operacional da deep fake foi um vídeo real do CEO pronunciando sobre a interferência russa no período eleitoral dos Estados Unidos (Kaspersky).

No Brasil, em 2022 os jornalistas Willim Bonner e Renata Vasconcellos foram vítimas desta prática tecnológica. Em tese, o vídeo manipulado era de ambos profissionais que relataram sobre os dados de pesquisa referente à intenção de votos a candidatos de preferência. O que chamou atenção no conteúdo era que os números de votos estavam trocados, relatado falsamente o candidato favorito (Schmidt, 2022).

O que se percebe nestes relatos e números é que apesar de conceder a liberdade de expressão e manifestação, as normas regulatórias precisam resguardar a verdadeira democracia. O emprego de deep fake impulsiona a polarização na internet, fazendo com que os indivíduos absorvam informações que lhe são agradáveis aos seus valores religiosos e de senso comum. A repercussão a longo prazo desse cenário é uma certa desconfiança sobre as informações divulgadas nas plataformas digitais, inclusive aquelas oriundas de entes governamentais. Deve-se impedir a dissipação massiva de informações falsas manipuladas pela tecnologia para fins de convencimento e prejuízo à terceiro (Schmidt, 2022).

No quesito regulatório, é possível enumerar dois principais projetos de lei propostos para cuidar sobre o tema. O PL nº 145/2024, cuja autoria é do Senador Chico Rodrigues, busca adaptar o Código de Defesa do Consumidor para tratar sobre o emprego das novas tecnologias na realização de propagandas. Neste caso, o objetivo é impedir que essas ferramentas sejam empregadas de forma maliciosa para prejudicar o consumidor final. O que chama atenção é que

o projeto tem como principal justificativa a presença da técnica de deep fake, o qual é capaz de influenciar negativamente vários setores da sociedade. Para solidificar a proposta, o texto traz as ameaças desta habilidade aos processos eleitorais, abordando como foco o consumidor e a sua vulnerabilidade diante das tecnologias de inteligência artificial na transmissão de informação e propagandas comerciais. A proposta então cuida em alterar o artigo 2º da Lei nº 8.078/1990 proibindo o uso de abordagens publicitárias manipuladas negativamente por ferramentas tecnológicas (Brasil, 2024).

O segundo PL é o nº 146/2024, também proposto pelo Senador Chico Rodrigues, mas com os esforços empenhados para alterar o Código Penal em duas situações. A primeira é atribuir como causa de aumento de pena o uso de deep fake em crimes contra a honra, enquanto a segunda é qualificar quando essas mesmas ferramentas são usadas no crime de falsa identidade. Da mesma forma que houve no PL nº 145, aqui o referido termo foi empregado para embasar a justificativa, de modo que em ambos os crimes há a prática comum em usar tecnologias de inteligência artificial para obter sucesso na empreitada criminosa. Para isso, o breve texto propõe uma alteração nos artigos 141 e 307 do Decreto Lei nº 2.848 de 1940 (Brasil, 2024).

Nos Estados Unidos já há sinais de regulações sobre o *deepfake* em alguns Estados, como na Califórnia. No Brasil, apesar das propostas legislativas, não há legislações que cuidam especificamente do tema, usando para isso abordagens análogas de outros textos legais (Berenguel, Molina, 2022, p. 6).

Desta forma, mesmo não havendo regulações sobre o *deepfake*, deve-se considerar que se trata de uma extensão das fakes News que está gradativamente presente na realidade das tecnologias de informação e comunicação. O compartilhamento de informações falsas solidifica o ambiente de desinformação e por consequência prejudica o repasse do conteúdo na sua forma íntegra. Esses fatores são amplificados pela polarização causada na internet e no desejo dos usuários em absorver apenas informações que lhe são palatáveis, independente de serem ou não verificadas.

O repasse dessas informações inverídicas é aperfeiçoado pelo deep fake, que com tecnologias de inteligência artificial, estas ficam quase que impossíveis de distinguir com aquelas verdadeiras. Com essa técnica digital, as notícias e seus respectivos conteúdos ficam ameaçados, especialmente aquelas que derivam de órgãos estatais. Por consequência, gera-se uma desconfiança de usuários na internet por parte dessas entidades públicas. Por isso, resta no próximo capítulo investigar um caso envolvendo um gestor Estadual vítima de deep fake e os prejuízos causados ao compartilhamento de informações íntegras oriundas da instituição

pública. Essa análise permitirá identificar a viabilidade de possíveis soluções, sejam iniciativas privadas ou políticas públicas.

3 ESTUDO DO CASO DO GOVERNADOR TARCÍSIO DE FREITAS E SOLUÇÕES AO DEEP FAKE

O aperfeiçoamento das habilidades de *deepfake* gera um impacto direto na disseminação de fake News. O resultado é um combate árduo das movimentações governamentais contra as barreiras à integridade da informação. Com um olhar mais recente, a agência Lupa realizou uma pesquisa nos meses de setembro e outubro de 2023, identificando imagens realizadas com *deepfake* na plataforma do Facebook. O resultado consistiu na verificação de pelo menos 30 conteúdos criados com esta técnica, sendo a maioria com imagens de indivíduos políticos. Dentre os rostos estavam Fernando Haddad, Jair Bolsonaro e Tarcísio de Freitas (Fagundes, 2024).

Isso justifica o fato de o tema integridade informacional ser inerido na agenda de discussões do G20. Por esta razão, caberá neste capítulo analisar o caso específico do governador Tarcísio de Freitas, vítima de *deepfake*, Em seguida haverá a necessidade de mapear algumas soluções contra o uso dessa tecnologia que é obstáculo do compartilhamento de informações verídicas.

O caso ocorreu no mês de setembro de 2024, quando o Governo do Estado de São Paulo emitiu um alerta referente ao compartilhamento massivo de um vídeo realizado com a mesma face e voz do Governador de São Paulo Tarcisio de Freitas do partido Republicanos. Após algumas investigações constatou-se que se tratava de um vídeo falso criado com o uso da inteligência artificial. O conteúdo consistia no pronunciamento do gestor em que o Procon estaria aplicando multas a empresas de cartão de crédito. No vídeo, havia a imposição da obrigação de fazer para que as operadoras restituíssem o valor pré-determinado em quantia monetária aos consumidores, sendo basicamente um processo de *cashback* (Carta Capital, 2024).

Ainda, o falso Tarcísio direciona os consumidores envolvidos a entrar no link disponível que encaminhava direito ao navegador falsificado do Procon. No acesso final, havia campos de preenchimento cujas informações solicitadas eram dados bancários e dados pessoais para que a restituição fosse possível. Acontece que o resultado final era o fornecimento de dados pessoais e dados pessoais sensíveis à golpistas autores do vídeo criado com *deepfake*. No pronunciamento realizado pelo Estado paulista, informou-se que “O Procon-SP e outros órgãos

do Governo de SP não promovem campanhas de cashback de compras de cartão de crédito" (Carta Capital, 2024).

No mesmo dia, um inquérito foi aberto e segue em investigação buscando o responsável pela autoria. Nota-se que neste caso há pelo menos cinco perspectivas desafiadoras: o ceticismo criado sobre a imagem de uma autoridade pública, a fraude cibernética envolvendo habilidades de fake News; às indenizações pelo vazamento de dados; a integridade da informação e por consequência a desconfiança posterior criada aos usuários da internet no que se refere à pronunciamentos de gestores públicos.

Mesmo com essas quatro problemáticas, nota-se que em razão da carência legal sobre o tema, há um caminho penoso para resguardar todas as vítimas envolvidas. Mesmo assim é possível viabilizar algumas alternativas que apresentam como solução neste cenário cibernético. Em primeiro lugar, pode-se sinalizar a empresa Sensity que desenvolveu softwares capazes de detectar imagens, vídeos e áudios manipulados tecnologicamente. Segundo as informações de seu próprio navegador, nos últimos 12 meses, mais de 35 mil *deepfake* foram identificados pelos seus próprios aplicativos, cuja análise é em média de 14 minutos por conteúdo (Sensity).

Seguindo esse mesmo raciocínio, a Google formulou um banco de dados formado por mais de 3 mil vídeos articulados com *deepfak*e para servir de matéria prima aos desenvolvedores de softwares detectores (Cancelier *apud* Dias *et al*, 2024, p. 17). Em termos acadêmicos, a *University at Buffalo Media Forensics Lav* desenvolveu uma plataforma de código aberto para identificar algoritmos derivados de aprendizado de máquina decorrentes das Redes Neurais Generativas Adversárias (University at Buffalo).

Em terceiro lugar, há ainda uma alternativa brasileira materializada como uma política pública da União. Consiste no chamado "Guia Ilustrativo Contra as *deepfakes*", elaborado pelo Supremo Tribunal Federal junto com a Data Privacy Brasil e publicado em 2024. No texto apresentado, há uma abordagem dinâmica sobre o conceito de *deepfake*, suas formas de ocorrência e meios de denúncia. O que chama atenção é que o documento cuida detalhando as formas de identificar uma imagem, vídeo ou áudio manipulado. Com exceção à *deepfakes* mais elaboradas, o guia menciona que naquelas mais comuns há pelo menos sete variações em vídeos que facilitam a identificação da articulação tecnológica (Supremo Tribunal Federal, 2024).

Essa iniciativa vem da política pública federal que é o Programa de Combate à Desinformação do STF, regulado pela Resolução nº 742, de 2021. O objetivo dessa ação é justamente combater as barreiras existentes por ferramentas tecnológicas na batalha contra o ambiente de desinformação. Deve-se notar que esta ação pública sinaliza preocupações já

abordadas neste estudo como confiança excessiva nas informações na internet e a ciberpolarização, sendo ambas ameaçadoras da formação crítica do cidadão brasileiro e por consequência do exercício da democracia. Outro ponto problemático abordado pelo programa é a formação a longo prazo da desconfiança à dados e relatos compartilhados por entes públicos e seus gestores. No caso do Governador Tarcísio, houve um efeito cascata do indivíduo que se encontra no ambiente hostil a informações verídicas e, por este motivo, atribui um ceticismo em pronunciamentos posteriores de ações governamentais.

Por isso, as estratégias do Programa são direcionadas a três principais eixos: "compreender a desinformação, reduzir o impacto das narrativas desinformativas, e recuperar a confiança das pessoas". Deste modo, os estudos são realizados para identificar as fases de construção de um cenário de desinformação. Em seguida formaliza-se parcerias com a sociedade, o terceiro setor e entidades privadas, para solidificar outras políticas públicas capazes de combater a disseminação de informações manipuladas. Por fim, o resultado é a materialização do terceiro eixo, que é a reconstrução da confiança popular sobre as informações compartilhadas na internet e a proteção da integridade de seus conteúdos (Supremo Tribunal Federal).

Exposta essa política pública protagonista no combate de fake news e deep fake, há uma quarta alternativa que está diretamente relacionada às iniciativas brasileiras mencionadas anteriormente. Compreende na alfabetização digital que é "ensinar o usuário da internet a usá-la com responsabilidade" (Jacob, 2020, p. 298). O Marco Civil da Internet, Lei nº 12.965/2014, indica em seu artigo 27, inciso II a necessidade de políticas públicas, fomentadoras do ambiente tecnológico, em garantir a inclusão digital. Em sintonia, o artigo 29, § Único da mesma lei prevê que os órgãos estatais devem viabilizar meios educacionais sobre o uso consciente da internet e as suas informações disponíveis (Brasil, 2014). Nesta linha de raciocínio, a inclusão digital refere-se em ter à disposição tecnologias de informação e comunicação, mas também condições para saber manusear essas ferramentas e o que elas oferecem (Fachin *et al*, 2022). É neste contexto que a alfabetização informacional insere como alternativa de solução ao combate à deep fake frente à integridade da informação, pois compreende em abordagens direcionadas à usuários da internet para que possam absorver a quantidade massiva de conteúdos e conseguí-los filtrá-los devidamente.

A quinta e última alternativa que merece ser exposta neste estudo são alguns princípios globais indicados pela Organização das Nações Unidas (ONU) para fortalecer a integridade da informação. Consiste em uma carta de Recomendações para Ações de Múltiplas Partes interessadas elaborada pela própria ONU e aderido pelo Brasil no ano de 2024. A ideia é que

estes princípios sejam observados na elaboração de políticas públicas nacionais atendendo as demandas problemáticas sinalizadas pelos números científicos mencionados anteriormente. Em primeiro lugar há o princípio da "confiança e resiliência social" (ONU, 2024), que aduz na compreensão de que todos os usuários da internet são frágeis. Logo há a necessidade de solidarizar-se com os grupos vulneráveis à as ameaças da desinformação e buscar por ações capazes de captar todas as linguagens informacionais, inclusive aquelas oriundas de sistemas computacionais.

Em segundo lugar há o princípio de "incentivos saudáveis" (ONU, 2024), traduzindo na importância de modelos de negócio buscarem o lucro e o desenvolvimento econômico a partir de procedimentos operacionais saudáveis, sem manipulações e informações falsas usadas para fins de controle sobre os seus consumidores. Já a "captação pública" é o terceiro princípio, que enfatiza a importância de os usuários terem o controle sobre o conteúdo disponível na internet e especialmente sobre sua natureza, ou seja, verídica ou não.

Os meios de comunicação independentes, livres e plurais" (ONU, 2024) confirmam a noção contrária ao caso em análise deste capítulo. Neste princípio há uma defesa em navegar pela internet com o uso de ferramentas tecnológicas sem estar vulnerável a possibilidades de se expor a conteúdos falsificados e dados pessoais vazados. Por fim, o último princípio é da "Transparência e pesquisa" (ONU, 2024) que comunica a importância de as ações públicas exigirem uma transparência nas tecnologias de informação e comunicação e nos veículos de acesso à internet. Além disso, impõe a importância de estudos científicos, como este que está sendo elaborado, sobre os riscos das tecnologias na garantia da integridade da informação e as possíveis soluções mitigadoras.

Em resumo, o caso de deep fake envolvendo o governador Tarcísio de Freitas sinaliza as ameaças sobre vários enfoques, como a proteção de dados, crime de fraude e especialmente à integridade da informação. Este último termo é permeado por contextos de fake news e desinformação, dos quais ambos são amplificados pelo aumento das novas tecnologias, inserindo, portanto, o termo *deepfake*. São informações manipuladas presentes na internet que impulsionam a ciberpolarização, deixando a democracia na sua forma mais frágil. No Brasil, apesar de ainda não haver regulações específicas sobre esta forma de manobrar um conteúdo com ferramentas de inteligência artificial, há alternativas materializadas em políticas públicas que viabilizam o combate do *deepfake* em face da integridade informacional.

CONCLUSÃO

A presente pesquisa teve como objetivo abordar noções gerais e variações de termos que contornam o conceito de integridade informacional e mapear os obstáculos criados pelo deepfake, com plano de fundo o caso do Governador do Estado de São Paulo Tarcisio de Freiras do partido Republicanos, que foi vítima da respectiva habilidade tecnológica.

Dentre as hipóteses verificadas, estão: a massificação de informações disponíveis na internet influenciadora da cultura da pós-verdade e da infodemia afetam diretamente no combate da integridade da informação; a criação do ambiente da desinformação e do ceticismo das relatos disponíveis na internet são intensificados com *deepfake*; e se ações governamentais, políticas públicas e iniciativas internacionais encontradas no estudo se materializam como possíveis soluções.

Inicialmente cuidou-se em abordar sobre a integridade informacional e o impacto das Tecnologias da Informação e Comunicação na disseminação em massa de informações compartilhadas na internet. Neste primeiro momento cuidou-se em compreender como os países, incluindo o Brasil, estão lidando com a nova realidade de fake news e a cultura da pós-verdade responsável por fomentar a ciberpolarização e o ambiente da desinformação.

Em seguida, analisou-se sobre o uso negativo das ferramentas de inteligência artificial na manipulação de imagens, vídeos e áudio, sendo esta técnica denominada em *deepfake*. Com isso, neste segundo momento analisou o funcionamento dessa habilidade tecnológica e como ela impacta diretamente no processo garantidor da integridade da informação. Destaca-se ainda relatos nacionais e internacionais de *deepfake*, além de possíveis regulações para cuidar especificamente sobre o tema.

No capítulo final, houve um estudo de caso do Governador do Estado de São Paulo, Tarcisio de Freitas do partido Republicanos, o qual foi vítima da prática de *deepfake*. Para tanto, este estudo identificou as problemáticas materializadas com o caso ocorrido, incluindo a integridade da informação, o vazamento de dados, o ceticismo sobre o perfil de agentes políticos e a posterior desconfiança dos usuários da internet em relação a pronunciamentos estatais.

Teve como pergunta central quais as estratégias e iniciativas públicas e privadas podem ser efetivadas e exteriorizadas em face dos desafios encontrados nos casos de deepfake ameaçadores da integridade da informação. Os principais impasses encontrados foram a falsa liberdade de manifestação e expressão, censura democrática, somada ao enraizamento da pós-verdade e a desinformação. São efeitos que prejudicam o acesso à informação verdadeira e o processo de identificação de autenticidade dos conteúdos disponíveis.

A pesquisa conseguiu atender os objetivos pontuando os planos estratégicos de entidades privadas e públicas essenciais para mitigar os riscos do *deepfake*, especialmente por ainda ser um tema não regulamentado pelas normas brasileiras. Com isso, foi objeto de discussão como as ameaças à integridade da informação causam impactos negativos no exercício da democracia e a real liberdade de expressão.

Deste modo, conclui-se que o efeito do *deepfake* gera impactos negativos diretos na garantia da integridade da informação e no estímulo da desinformação. Apesar do Brasil não contar com legislações específicas sobre o tema, é possível identificar ações estatais e privadas para mitigar os riscos verificados na pesquisa. As discussões promovidas neste estudo permitem que estes planos e políticas públicas possam ser efetivados e aperfeiçoados por gestores e legisladores sobre o tema.

REFERÊNCIAS

AGÊNCIA ESTADO. 'Fake news' se espalham 70% mais rápido que notícias verdadeiras, diz MIT. **Correio Braziliense**, 08 mar. 2018. Disponível em: https://www.correiobraziliense.com.br/app/noticia/tecnologia/2018/03/08/interna_tecnologia,664835/fake-news-se-espalham-70-mais-rapido-que-noticias-verdadeiras.shtml. Acesso em 14 mar. 2025.

AGUIAR, Alexandre Kehring Veronese; FACHIN, Jéssica Amanda; LANNES, Yuri Nathan da Costa. Políticas Públicas De Acesso E Universalização Da Internet No Brasil E Cidadania Digital. **Revista de Direito Brasileira**. Florianópolis, v. 32, n. 12, p. 110-129, mai/ ago. 2022

ARAÚJO, Carlos Alberto Ávila. Integridade da informação: um possível novo conceito para o estudo da desinformação. **Revista Comunicação Midiática,** Bauru, SP, v. 19, n. 1, p. 207-226, 2024. DOI: https://doi.org/10.5016/gpkkyf59. Disponível em: https://www2.faac.unesp.br/comunicacaomidiatica/index.php/CM/article/view/614. Acesso em 14 mar. 2025

ARAÚJO, Marilene, SANTOS, Maria Celeste. Cordeiro Leite dos. Integridade da informação: interfaces entre direito e inteligência artificial C20 2024/ G20/WG7 – Digitalização e Tecnologia. **Revista Internacional Consinter de Direito**, n. 10, p. 843-873, dez. 2024. DOI: 10.19135/revista.consinter.00019.40. Disponível em: https://revistaconsinter.com/index.php/ojs/article/view/734. Acesso em 14 mar. 2025

BERENGUEL, Orlando Leonardo. MOLINA, Adriano Cezar. Integridade da informação: interfaces entre direito e inteligência artificial C20 2024/ G20/WG7 – Digitalização e Tecnologia. **Revista Sociedade e Desenvolvimento**. São Paulo, v. 11, n. 6, p. 1-9, mai. 2022. DOI: https://doi.org/10.33448/rsd-v11i6.29533. Disponível em: https://rsdjournal.org/index.php/rsd/article/view/29533. Acesso em 14 mar. 2025

BRASIL, Câmara dos Deputados. Projeto de Lei nº 145, de 06 de fev. de 2024. Altera a Lei nº 8.078, de 11 de setembro de 1990 (Código de Defesa do Consumidor), para regular o uso de ferramentas de inteligência artificial para fins publicitários e coibir a publicidade enganosa com uso dessas ferramentas. Brasília: Câmara dos Deputados, 2024. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/161946. Acesso em 14 mar. 2025.

BRASIL, Câmara dos Deputados. Projeto de Lei nº 146, de 06 de fev. de 2024. Altera o Decreto-Lei nº 2.848, de 7 de dezembro de 1940 (Código Penal), para estabelecer causa de aumento de pena para os crimes contra a honra e hipótese qualificada para o crime de falsa identidade, para quando houver a utilização de tecnologia de inteligência artificial para alterar a imagem de pessoa ou de som humano. Brasília: Câmara dos Deputados, 2024. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/161947. Acesso em 14 mar. 2025.

CARLUCCI, Manoela. Governo de SP alerta para vídeo fraudulento que usa de deep fake para recriar voz e imagem de Tarcísio. **CNN Brasil**, 13 set. 2024. Disponível em: https://www.cnnbrasil.com.br/politica/governo-de-sp-alerta-para-video-fraudulento-que-usa-de-deep-fake-para-recriar-voz-e-imagem-de-tarcisio/. Acesso em 14 mar. 2025.

CARTA CAPITAL. **O que se sabe sobre o golpe que usa imagem e voz de Tarcísio de Freitas geradas por IA**. 13 set. 2024. Disponível em: https://www.cartacapital.com.br/cartaexpressa/o-que-se-sabe-sobre-o-golpe-que-usa-imagem-e-voz-de-tarcisio-de-freitas-geradas-por-ia/ Acesso em 14 mar. 2025

CETIR.br. TIC Domicílios – 2024: A1 – Usuários de internet, por tipo de habilidade digital. São Paulo: Cetic.br, [s.d.]. Disponível em: https://cetic.br/pt/tics/domicilios/2024/individuos/I1A/. Acesso em: 30 jun. 2024.

COELHO, Thomaz. Golpistas usam IA para falsificar voz do governador Tarcísio e roubar dados. **CNN Brasil**, 13 set. 2024. Disponível em: https://www.cnnbrasil.com.br/nacional/golpistas-usam-ia-para-falsificar-voz-do-governador-tarcisio-e-roubar-dados/. Acesso em 14 mar. 2025.

DIAS, José Wanderley Dallas Reis. GATINHO, Gislaine Fernanda Carvalho. SILVEIRA, Túlio Belchior Mano da. Consequências jurídicas da ciberpolarização do deepfake face ao estado democrático de direito. **Revista Contribuciones a Las Ciencias Sociales,** v.17, n. 10, p. 01-23, out. 2024. Disponível em: https://ojs.revistacontribuciones.com/ojs/index.php/clcs/article/view/12175#:~:text=Assim%2C%20o%20fen%C3%B4meno%20da%20ciberpolariza%C3%A7%C3%A3o,a%20prote%C3%A7%C3%A3o%20da%20dignidade%20humana. Acesso em 14 mar. 2025

FAGUNDES, Evelyn. Deepfakes de políticos para aplicar golpes avançam nas redes; Marçal e Bolsonaro são principais alvos, **LUPA,** 06 nov. 2024. Disponível em: https://lupa.uol.com.br/jornalismo/2024/11/06/deepfakes-de-politicos-para-aplicar-golpes-avancam-nas-redes-marcal-e-bolsonaro-sao-principais-alvos. Acesso em 14 mar. 2025

GUIMARÃES, Pedro. RODRIGUES, Cleber. 4 em cada 10 brasileiros afirmam receber fake news diariamente. **CNN Brasil**, 29 ago. 2022. Disponível em:

https://www.cnnbrasil.com.br/nacional/4-em-cada-10-brasileiros-afirmam-receber-fake-news-diariamente/. Acesso em 14 mar. 2025.

INTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Ibict promove 'Conferência Livre: Ciência no Combate à Desinformação' na CAPES**. 26 mar. 2024. Disponível em: https://www.gov.br/ibict/pt-br/central-de-conteudos/noticias/2024/marco/ibict-promove-2018conferencia-livre-ciencia-no-combate-a-desinformacao2019-na-capes. Acesso em 14 mar. 2024

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Pesquisa Nacional Por Amostra de Domicílios Contínua**. Rio de Janeiro, RJ. ISBN: 9788524046063. 16 p. 2024. Disponível em: https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2102068. Acesso em 14 mar. 2025

JACOB, Paola Domingues. Análise do pacto democrático no universo das fake news no Brasil. **Revista RH Visão Sustentável,** Rio de Janeiro, v. 2, n. 4, p. 288-303, jul./ dez.2020. Disponível em: https://revistas.cesgranrio.org.br/index.php/rh_visaosustentavel/article/view/3275/1405. Acesso em 14 mar. 2025

JARDIM, José Maria. A LEI DE ACESSO À INFORMAÇÃO PÚBLICA: dimensões político-informacionais. **Revista Tendências da Pesquisa Científica em Ciência da Informação.** Espírito Santo, v. 5, n. 1, p. 1-22, jan. 2012. Disponível em: https://revistas.ancib.org/index.php/tpbci/article/view/266. Acesso em 14 mar. 2025.

KASPERSKY TEAM. **Pesquisa: a infodemia e os impactos na vida digital**. 11 mai. 2021. Disponível em: https://www.kaspersky.com.br/blog/pesquisa-infodemia-impactos-vida-digital/17467/. Acesso em 14 mar. 2025

KASPERSKY TEAM. **Vídeos falsos e deepfake – Como os usuários podem se proteger**. Disponível em: https://www.kaspersky.com.br/resource-center/threats/protect-yourself-from-deep-fake. Acesso em 14 mar. 2025

LYU, Siwei. DeepFake-o-meter: exposing DeepFakes. **University at Buffalo**. Disponível em: https://www.buffalo.edu/digital-scholarship-studio-network/projects/faculty-projects/DeepFake-o-meter.html. Acesso em 14 mar. 2025.

NÚCLEO DE INFORMAÇÃO E COORDENAÇÃO DO PONTO BR. **Pesquisa revela que 43% dos jovens não sabem checar se uma informação da Internet é falsa**. 03 mai. 2023. Disponível em: https://www.nic.br/noticia/na-midia/pesquisa-revela-que-43-dos-jovens-nao-sabem-checar-se-uma-informacao-da-internet-e-falsa/. Acesso em 14 mar. 2025

MONTORO, Ana Carolina. Voz de Tarcísio é recriada com IA para aplicar fraude nas redes sociais. **Exame**, 13 set. 2024. Disponível em: https://exame.com/brasil/voz-de-tarcisio-de-freitas-e-recriada-por-inteligencia-artificial-para-aplicar-fraudes/. Acesso em 14 mar. 2025.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. **Princípios Globais das Nações Unidas para a Integridade da Informação - Recomendações para Ação de Múltiplas Partes Interessadas**. Paris: 2024. Disponível em: https://brasil.un.org/pt-br/274644-

princ%C3%ADpios-globais-para-integridade-da-informa%C3%A7%C3%A3o. Acesso em 14 mar. 2025.

PSafe PROTEGE. **Pesquisa fake news no Brasil em 2020**. 2020. Disponível em: https://www.psafe.com/blog/wp-content/uploads/2020/10/Pesquisa-Fake-News-PSafe-2020.pdf. Acesso em 14 mar. 2025

SENSITY AI. **The Sate Of Deepfake 2024**. 2024. Disponível em: https://5865987.fs1.hubspotusercontent-na1.net/hubfs/5865987/SODF%202024.pdf. Acesso em 14 mar. 2025

SECRETARIA DE COMUNICAÇÃ SOCIAL. **Brasil lidera diálogos sobre integridade da informação e regulação de plataformas**. 02 mai. 2024. Disponível em: https://www.gov.br/secom/pt-br/assuntos/noticias/2024/05/brasil-lidera-dialogos-sobre-integridade-da-informacao-e-regulacao-de-plataformas. Acesso em 14 mar. 2025

SECRETARIA DE COMUNICAÇÃ SOCIAL. **Brasil adere à Recomendação da OCDE sobre Integridade da Informação**. 17 dez. 2024. Disponível em: https://www.gov.br/secom/pt-br/assuntos/noticias/2024/12/brasil-adere-a-recomendacao-da-ocde-sobre-integridade-da-informacao. Acesso em 14 mar. 2024

SCHMIDT, Sarah. Deepfakes, o novo estágio tecnológico da desinformação Algoritmo detecta imagens e vídeos alterados com inteligência artificial. **Pesquisa FAPESP**, São Paulo, e. 321, 20 dez. 2022. Disponível em: https://revistapesquisa.fapesp.br/deepfakes-o-novo-estagio-tecnologico-das-noticias-falsas/. Acesso em 14 mar. 2025.

SUPREMO TRIBUNAL FEDERAL. **Programa de Combate à Desinformação: sociedade informada, democracia forte**. Disponível em: https://portal.stf.jus.br/desinformacao/. Acesso em 14 mar. 2024

TORRES, Lívia. Pesquisa aponta que WhatsApp é a principal fonte de informação de 79% dos entrevistados. **Rádio Senado**, 12 dez. 2019. Disponível em: https://www12.senado.leg.br/radio/1/noticia/2019/12/12/pesquisa-aponta-que-whatsapp-e-a-principal-fonte-de-informacao-de-79-dos-entrevistados. Acesso em: 4 fev. 2019.

TRIBUNAL REGIONAL ELEITORAL-GO. **Fake news x desinformação: entenda qual é a diferença entre os termos**. 23 ago. 2023. Disponível em: https://www.tre-go.jus.br/comunicacao/noticias/2023/Agosto/fake-news-x-desinformacao-entenda-qual-e-a-diferenca-entre-os-termos. Acesso em 14 mar. 2024