FACULDADE DE TECNOLOGIA DE AMERICANA "MINISTRO RALPH BIASI"
Curso Superior de Tecnologia em Segurança da Informação

Edmar Luiz Lucas

Manipulação Digital em Rede: *Deepfakes*, *Bots* e Milícias Virtuais como Desafios Contemporâneos à Segurança da Informação.

Americana, SP
2025

FACULDADE DE TECNOLOGIA DE AMERICANA "MINISTRO RALPH BIASI"

Curso Superior de Tecnologia em Segurança da Informação

Edmar Luiz Lucas

Manipulação Digital em Rede: *Deepfakes*, *Bots* e Milícias Virtuais como Desafios Contemporâneos à Segurança da Informação.

Projeto monográfico, desenvolvido em cumprimento à exigência curricular do Curso Superior de Tecnologia em Segurança da Informação da Fatec Americana, sob orientação da Professora Esp. Ana Lucia Spigolon..

Área de concentração: Manipulação Digital em Redes de Computador

Americana, SP.

2025

FICHA CATALOGRÁFICA – Biblioteca Fatec Americana Ministro Ralph Biasi- CEETEPS Dados Internacionais de Catalogação-na-fonte

LUCAS, Edmar L.

Manipulação Digital em Rede: Deepfakes, Bots e Milícias Virtuais como Desafios Contemporâneos à Segurança da Informação. / Edmar L. Lucas – Americana, 2025.

33f.

Monografia (Curso Superior de Tecnologia em Segurança da Informação) - - Faculdade de Tecnologia de Americana Ministro Ralph Biasi – Centro Estadual de Educação Tecnológica Paula Souza

Orientadora: Profa. Esp. Ana Lúcia Spigolon

1. Segurança em sistemas de informação. I. LUCAS, Edmar L. II. SPIGOLON, Ana Lúcia III. Centro Estadual de Educação Tecnológica Paula Souza – Faculdade de Tecnologia de Americana Ministro Ralph Biasi

CDU: 681.518.5

Elaborada pelo autor por meio de sistema automático gerador de ficha catalográfica da Fatec de Americana Ministro Ralph Biasi.

Edmar Luiz Lucas

Análise de ameaças em emails: uma avaliação comparativa entre julgamento humano e inteligência artificial

Trabalho de graduação apresentado como exigência parcial para obtenção do título de Tecnólogo em Curso Superior de Tecnologia em Segurança da Informação pelo Centro Paula Souza – FATEC Faculdade de Tecnologia de Americana Ministro Ralph Biasi.
Área de concentração: Segurança da informação.

Americana, 01 de dezembro de 2025.

Banca Examinadora:

Ana Lúcia Spigolon
Especialista
Fatec Americana "Ministro Ralph Biasi"

Jonas Bodê
Mestre
Fatec Americana "Ministro Ralph Biasi"

Paula da Fonte Sanches
Mestre
Fatec Americana "Ministro Ralph Biasi"

AGRADECIMENTOS

Agradeço profundamente à minha esposa e aos meus filhos e netos, pelo apoio incondicional e pela compreensão ao longo desta jornada. Agradeço também aos amigos, que estiveram ao meu lado, e aos professores, cuja paciência e dedicação em compartilhar seu saber foram fundamentais para a construção do conhecimento que hoje possuo. Sem o incentivo e o suporte de todos eles, este trabalho não teria sido possível.

DEDICATÓRIA

Dedico este trabalho à minha esposa e aos meus filhos, cuja paciência, amor e compreensão foram fontes de força e inspiração ao longo de cada etapa desta trajetória. A eles, que me motivam diariamente a buscar o melhor de mim e nunca desistir dos meus objetivos. Dedico também aos professores e amigos, que compartilharam seu conhecimento e apoio com generosidade e entusiasmo. Que este trabalho seja uma pequena retribuição à confiança e ao incentivo que sempre depositaram em mim.

RESUMO

A expansão acelerada das tecnologias de inteligência artificial transformou profundamente a forma como indivíduos interagem, se informam e constroem percepções no ambiente digital. Nesse contexto, práticas como *deepfakes*, *bots* inteligentes e milícias digitais deixaram de ser fenômenos isolados para se tornarem parte de um cenário cotidiano, no qual a fronteira entre o que é autêntico e o que é fabricado torna-se cada vez mais difícil de distinguir. Este estudo busca compreender, de maneira crítica e sensível, como essas formas de manipulação exploram fragilidades humanas — como confiança, curiosidade e urgência — e como se aproveitam da velocidade das redes sociais para produzir impactos reais na vida das pessoas, nas instituições e nas relações sociais. A pesquisa analisa casos documentados no Brasil e no exterior, evidenciando que a ação coordenada de conteúdos sintéticos compromete não apenas a segurança da informação, mas também valores democráticos e direitos fundamentais. Examina-se, ainda, o arcabouço jurídico brasileiro, que, embora disponha de instrumentos importantes, revela-se insuficiente frente à sofisticação dessas novas ameaças. A partir dessa leitura integrada, o trabalho evidencia que enfrentar a manipulação digital não depende apenas de mecanismos tecnológicos ou de punições legais, mas de um conjunto articulado de estratégias que inclui educação midiática, transparência algorítmica, cultura de prevenção e fortalecimento da capacidade crítica dos usuários. Conclui-se que a construção de um ambiente digital mais seguro exige não apenas normas e ferramentas, mas também um olhar humano atento às vulnerabilidades emocionais, sociais e informacionais que tornam possível a expansão desses fenômenos, reforçando a necessidade de respostas colaborativas e de consciência coletiva para preservar a integridade do ecossistema digital.

Palavras Chave: Segurança da Informação; Engenharia Social; Inteligência Artificial; *Deepfake*; Milícias Digitais..

ABSTRACT

The rapid expansion of artificial intelligence technologies has profoundly transformed the way individuals interact, access information, and form perceptions in digital environments. In this context, practices such as deepfakes, intelligent bots, and digital militias have ceased to be isolated phenomena and have become part of a daily landscape in which the boundary between what is authentic and what is fabricated grows increasingly difficult to discern. This study seeks to understand, through a critical and human-centered perspective, how these forms of manipulation exploit emotional vulnerabilities—such as trust, curiosity, and urgency—and how they take advantage of the speed and reach of social networks to generate concrete impacts on people's lives, institutions, and social relations. The research examines documented cases in Brazil and abroad, showing that the coordinated action of synthetic content compromises not only information security but also democratic values and fundamental rights. It also analyzes the Brazilian legal framework, which, although it offers relevant instruments, remains insufficient in the face of the sophistication of these emerging threats. From this integrated perspective, the study demonstrates that confronting digital manipulation depends not only on technological mechanisms or legal sanctions but on an articulated set of strategies that include media literacy, algorithmic transparency, preventive culture, and the strengthening of users' critical awareness. The findings conclude that building a safer digital environment requires more than norms and tools; it demands a human perspective attentive to the emotional, social, and informational vulnerabilities that enable the expansion of these phenomena, reinforcing the need for collaborative responses and collective awareness to preserve the integrity of the digital ecosystem.

Keywords: Information Security; Social Engineering; Artificial Intelligence; Deepfake; Digital Militias.

LISTA DE ABREVIATURAS E SIGLAS

CEETEPS – Centro Estadual de Educação Tecnológica Paula Souza

CNJ – Conselho Nacional de Justiça

DAS – Digital Services Act (Lei dos Serviços Digitais – LSD)

ENISA – European Union Agency for Cybersecurity

FATEC – Faculdade de Tecnologia

GAN – Generative Adversarial Network, (Rede Generativa Adversarial)

IA – Inteligência Artificial

LGPD – Lei Geral de Proteção de Dados

PL – Projeto de Lei

STF – Supremo Tribunal Federal

TSE – Tribunal Superior Eleitoral

SUMÁRIO

1 INTRODUÇÃO

A manipulação digital em rede, impulsionada por técnicas avançadas de inteligência artificial, tornou-se uma das mais complexas ameaças contemporâneas à segurança da informação. *Deepfakes* capazes de substituir identidades, *bots* que simulam interação humana e milícias virtuais estruturadas para influenciar comportamentos coletivos formam um ecossistema de risco que opera em escala e velocidade inéditas. Nesse ambiente, a desinformação passa a ser produzida como produto tecnológico, e a percepção humana, antes barreira natural à fraude, transforma-se em vulnerabilidade explorável. Trata-se de um fenômeno que desafia não apenas os controles técnicos, mas a própria capacidade social de reconhecer o verdadeiro, o legítimo e o confiável.

Vivencia-se um tempo em que a fronteira entre o verdadeiro e o falso se torna cada vez mais tênue. Imagens são recriadas, vozes são clonadas e identidades inteiras podem ser fabricadas por algoritmos capazes de reproduzir com perfeição a aparência e o tom humano. A inteligência artificial, que nasceu como promessa de progresso, transforma-se também em instrumento de manipulação quando colocada a serviço de interesses escusos. Nesse cenário, a segurança da informação deixa de ser um tema restrito à tecnologia e passa a ocupar o centro do debate ético, social e jurídico contemporâneo.

As novas formas de engenharia social mediadas por inteligência artificial utilizam o conhecimento sobre o comportamento humano como principal arma. Não atacam servidores ou códigos, mas emoções: confiança, medo, curiosidade, empatia. *Deepfakes* que distorcem a imagem de pessoas, *bots* que espalham desinformação com aparência de legitimidade e redes coordenadas — as chamadas milícias digitais — moldam percepções coletivas e representam ameaças que transcendem o âmbito técnico, alcançando a própria noção de verdade e a estabilidade das instituições democráticas.

O objetivo deste trabalho consiste em compreender e analisar essas transformações. Pretende-se identificar como a inteligência artificial potencializa práticas de engenharia social, examinar de que modo o ordenamento jurídico brasileiro responde ao fenômeno e refletir sobre os desafios éticos e probatórios que

emergem quando o real e o sintético se confundem. Busca-se, ainda, propor medidas preventivas e educativas que fortaleçam a consciência crítica dos cidadãos e promovam uma cultura digital orientada pela responsabilidade e pela verdade.

A pesquisa fundamenta-se em revisão bibliográfica e documental de legislações, relatórios técnicos e estudos de caso internacionais, articulando fundamentos da segurança da informação com princípios de direitos fundamentais e responsabilidade social. A metodologia é qualitativa e exploratória, permitindo compreender as dimensões humanas que estruturam as ameaças digitais contemporâneas.

Estruturalmente, o trabalho organiza-se em seis capítulos: o primeiro refere-se a introdução deste trabalho; o segundo aborda os conceitos de segurança da informação e engenharia social; o terceiro examina o marco legal brasileiro e as discussões regulatórias internacionais; o quarto analisa casos emblemáticos de manipulação digital; o quinto propõe medidas jurídicas, tecnológicas e educacionais voltadas à prevenção; e o sexto reúne as considerações finais, reafirmando a necessidade de um olhar humano sobre a tecnologia e de uma ação coletiva em defesa da integridade informacional.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta as bases necessárias para entender como a manipulação digital acontece hoje. Primeiro, aborda o que é segurança da informação e por que ela vai além da tecnologia. Depois, explica como a engenharia social evoluiu e como a inteligência artificial passou a criar novos tipos de golpes e interações enganosas nas redes sociais.

2.1 Segurança da informação: conceitos e dimensões

A segurança da informação, conforme delineado pela ABNT NBR ISO/IEC 27000, compreende a preservação da confidencialidade, integridade e disponibilidade das informações, podendo incluir outros atributos, como autenticidade, responsabilidade e confiabilidade. A literatura técnica, representada por autores como Stallings (2017), enfatiza a relevância de mecanismos criptográficos, protocolos de autenticação e controles de proteção para mitigar riscos associados ao acesso, à modificação, à divulgação ou à destruição indevida de dados. Trata-se de um campo intrinsecamente interdisciplinar, que envolve dimensões técnicas, organizacionais, legais e comportamentais voltadas à garantia de que a informação — ativo estratégico de indivíduos, empresas e instituições — permaneça útil, íntegra e confiável ao longo do tempo

Tradicionalmente, a literatura especializada estrutura a segurança da informação em torno de três pilares: confidencialidade, integridade e disponibilidade. A confidencialidade assegura que apenas pessoas autorizadas possam acessar determinado conteúdo; a integridade garante que a informação permaneça fiel à sua forma original, sem alterações não autorizadas; e a disponibilidade refere-se à possibilidade de acesso às informações e aos sistemas sempre que necessário. Em modelos mais recentes, acrescenta-se a esses elementos a autenticidade, que assegura a veracidade da origem da informação, e a não repudiação, que impede a negação de autoria ou de participação em determinada ação (ABNT, 2020).

Embora a evolução tecnológica tenha ampliado os mecanismos de proteção técnica - como criptografia, autenticação multifator e sistemas de monitoramento -, as ameaças também se tornaram mais sofisticadas. Isso evidencia que a segurança da

informação não se limita à defesa contra falhas técnicas, mas deve contemplar estratégias de mitigação de riscos que exploram o comportamento humano, em especial nas interações mediadas por ambientes digitais (ENISA, 2023).

2.2 Engenharia social: conceitos e evolução

A engenharia social é definida como o conjunto de técnicas voltadas à manipulação psicológica de indivíduos, de modo a levá-los a revelar informações sigilosas ou a realizar ações que comprometm a segurança de sistemas, redes ou dados (Mitnick; Simon, 2003). Ao contrário de ataques puramente técnicos, que exploram vulnerabilidades de *software* ou *hardware*, a engenharia social direciona-se a vulnerabilidades cognitivas e emocionais, explorando a confiança, o medo, a curiosidade e até mesmo o senso de urgência das vítimas (Schmitt; Flechais, 2023).

Historicamente, as formas mais conhecidas de engenharia social incluem o *phishing* — envio de comunicações fraudulentas com o intuito de obter dados pessoais ou financeiros — e o *spear phishing*, que é o phishing direcionado a um alvo específico, com mensagens personalizadas. Outras variações incluem o *pretexting* (invenção de cenários para obtenção de informações), o *baiting* (oferta de benefícios falsos para induzir cliques ou downloads) e o *tailgating* (entrada física em áreas restritas acompanhando alguém autorizado) (MITNICK; SIMON, 2003).

Nos últimos anos, porém, observa-se a transição para modalidades de engenharia social mediadas por inteligência artificial e automação, capazes de gerar conteúdo altamente convincente e interações em escala, sem a necessidade de intervenção humana constante (Shu *et al*., 2022). Essa evolução amplia o alcance, reduz custos para os agentes maliciosos e eleva significativamente o grau de dificuldade na detecção das fraudes.

2.3 Inteligência artificial generativa e manipulação digital

Segundo Russell e Norvig (2013), os sistemas de inteligência artificial consistem em agentes capazes de perceber o ambiente, processar informações e tomar decisões a partir de dados, formando a base conceitual que sustenta os

modelos contemporâneos. No âmbito específico da inteligência artificial generativa, destacam-se arquiteturas como redes generativas adversariais (do Inglês: GANs), autoencoders variacionais (VAEs), modelos de difusão e transformadores, capazes de produzir novos conteúdos com elevado grau de realismo a partir de padrões aprendidos nos dados de treinamento. Estudos recentes demonstram que essas abordagens têm ampliado significativamente o potencial de geração de conteúdo sintético em diversos domínios, consolidando a IA generativa como um dos eixos centrais da inovação tecnológica contemporânea (BIBRI, 2025).

No campo da manipulação digital, a IA generativa viabiliza práticas antes restritas a especialistas, como a criação de *deepfakes*, clonagem de voz e geração de perfis falsos altamente verossímeis em redes sociais. Diferentemente de simples edições ou montagens, essas técnicas permitem simular expressões faciais, entonações e gestos, tornando a detecção extremamente desafiadora (Pawelec, 2022).

O risco à segurança da informação emerge quando tais recursos são empregados para criar narrativas falsas, fabricar provas digitais, manipular opiniões ou induzir comportamentos. O impacto potencial abrange desde prejuízos financeiros e danos reputacionais até a interferência em processos democráticos, como eleições, e a erosão da confiança pública em registros audiovisuais (Europol, 2024).

2.4 Novas formas de engenharia social em redes sociais

Entre as modalidades emergentes impulsionadas por IA e observadas nas redes sociais, destacam-se:

- *Deepfakes* em tempo real: transmissões de vídeo ao vivo ou reuniões virtuais nas quais a imagem e a voz do interlocutor são simuladas por sistemas de IA, permitindo fraudes complexas, como ordens falsas emitidas por supostos superiores hierárquicos (Cincodias, 2025).
- *Spam* social inteligente: uso de *bots* e contas falsas capazes de interagir de forma convincente com usuários reais, promovendo *links* maliciosos, desinformação ou campanhas de difamação (Techradar Pro, 2025).

- Milícias digitais: grupos organizados, frequentemente coordenados por algoritmos, que disseminam narrativas específicas, realizam ataques coordenados a alvos determinados e manipulam métricas de engajamento para amplificar conteúdos (STF, 2020).

Essas práticas não apenas desafiam mecanismos tradicionais de detecção, como também se beneficiam da velocidade e do alcance das redes sociais, potencializando o dano antes que medidas corretivas possam ser adotadas (Shu *et al.*, 2022).

3 ASPECTOS JURÍDICOS E REGULATÓRIOS

Este capítulo apresenta como o Direito brasileiro e internacional têm buscado responder aos desafios trazidos pela manipulação digital. Ao longo das próximas seções, são analisadas as principais normas que tratam de privacidade, proteção de dados e crimes cibernéticos, além de projetos de lei voltados especificamente às *fake news* e ao uso abusivo de tecnologias de inteligência artificial. Também são examinadas decisões judiciais recentes e exemplos de regulamentações estrangeiras, permitindo compreender como diferentes países vêm tentando equilibrar inovação, segurança e liberdade de expressão.

3.1 Marco legal da segurança da informação no Brasil

O arcabouço de jurídico brasileiro dispõe de um conjunto de normas que, embora não trate a segurança da informação de forma unificada, estabelece parâmetros relevantes para a proteção de dados, a responsabilização de agentes e a regulação do ambiente digital. Nesse cenário, o Tribunal Superior Eleitoral destaca que a consolidação de marcos regulatórios é elemento essencial para o enfrentamento de práticas desinformativas e para o fortalecimento da integridade do ecossistema informacional (TSE, 2022).

O Marco Civil da Internet (Lei nº 12.965/2014) institui princípios, garantias, direitos e deveres para o uso da rede no Brasil, assegurando, em seu artigo 3º, a proteção da privacidade, a neutralidade de rede e a segurança dos dados. O artigo 19, por sua vez, define as bases para a responsabilidade dos provedores de aplicações, determinando que a remoção de conteúdos ilícitos depende, como regra, de ordem judicial, com exceção dos casos de nudez ou ato sexual de caráter privado. Embora não tenha sido concebido especificamente para lidar com ameaças contemporâneas como deepfakes e milícias digitais, o Marco Civil fornece fundamentos interpretativos relevantes para discutir a atuação das plataformas e sua responsabilidade por conteúdos nocivos.

A Lei Geral de Proteção de Dados Pessoais – LGPD (Lei nº 13.709/2018) introduziu um regime abrangente de proteção à privacidade e ao tratamento de dados

pessoais, aplicável tanto ao setor público quanto ao privado. Ainda que voltada principalmente à coleta e ao uso de dados, a LGPD se relaciona diretamente com o tema deste estudo, pois conteúdos produzidos por inteligência artificial — como *deepfakes* — frequentemente se baseiam no uso não autorizado de imagens, vozes e demais informações pessoais. O artigo 46 impõe aos agentes de tratamento a adoção de medidas técnicas e administrativas aptas a proteger os dados contra acessos não autorizados, enquanto o artigo 42 estabelece deveres de reparação e responsabilidade civil.

Além disso, a Lei nº 14.155/2021 alterou o Código Penal para tipificar a fraude eletrônica, ampliando significativamente as penas quando a conduta envolve informações obtidas por meio de redes sociais, aplicativos de mensagens ou técnicas de phishing (Brasil, 2021). Embora ainda não alcance todas as formas emergentes de manipulação digital, a alteração legislativa representa um avanço relevante na responsabilização penal por crimes cibernéticos, sobretudo ao reconhecer a especificidade e a gravidade das infrações praticadas em ambiente digital.

3.2 Iniciativas legislativas específicas: o PL 2630/2020

O Projeto de Lei nº 2.630/2020, conhecido como Lei das *Fake News*, tramita no Congresso Nacional com o objetivo de estabelecer normas de transparência para provedores de redes sociais e serviços de mensageria privada (Brasil, 2020). Entre as medidas propostas estão a identificação de contas automatizadas (*bots*), a rastreabilidade de mensagens encaminhadas em massa e a obrigação de remoção célere de conteúdos ilícitos ou que violem direitos fundamentais.

Apesar das controvérsias sobre seus potenciais impactos na liberdade de expressão, o PL 2630/2020 busca enfrentar fenômenos como campanhas coordenadas de desinformação e a atuação de milícias digitais, que operam de forma articulada para influenciar a opinião pública e, em alguns casos, desestabilizar processos democráticos. Caso aprovado, poderá fornecer instrumentos jurídicos mais adequados para lidar com as práticas aqui estudadas, sobretudo no que diz respeito à responsabilização das plataformas e à transparência das interações automatizadas.

3.3 Precedentes judiciais e desafios probatórios

A jurisprudência brasileira tem começado a lidar com litígios envolvendo conteúdos sintéticos e campanhas de desinformação, embora ainda de forma incipiente. Em decisões recentes, tribunais têm determinado a remoção de *deepfakes* ofensivos, especialmente em contextos de violência contra a mulher e de assédio moral ou sexual, reconhecendo o uso indevido da imagem e a violação à honra (TJSP, Apelação nº 1004563-77.2021.8.26.0565, 2021).

No contexto eleitoral, o Tribunal Superior Eleitoral (TSE) já sinalizou preocupação com a disseminação de conteúdos manipulados, determinando, em casos específicos, a remoção de publicações e a aplicação de multas por propaganda irregular, com base na Resolução TSE nº 23.610/2019 (TSE, 2019). A dificuldade, contudo, reside na detecção e comprovação da falsidade do material, sobretudo diante de *deepfakes* de alta qualidade (Shu *et al*., 2022).

O desafio probatório é ampliado pela volatilidade dos conteúdos publicados em redes sociais, que podem ser apagados, alterados ou redistribuídos com extrema rapidez, além da atuação de contas automatizadas que dificultam a identificação das origens dos ataques. Nesse cenário, torna-se indispensável que vítimas e autoridades adotem medidas céleres de preservação da prova digital, recorrendo a instrumentos como a ata notarial prevista no artigo 384 do Código de Processo Civil, à captura de tela acompanhada de metadados verificáveis e ao registro de elementos técnicos que assegurem a autenticidade da evidência, como hash criptográfico, logs de acesso e informações de integridade fornecidas por plataformas. O próprio Conselho Nacional de Justiça tem reiterado a necessidade de adoção de boas práticas de cadeia de custódia digital, destacando que a preservação imediata é condição essencial para a confiabilidade da prova em procedimentos investigativos e judiciais (CNJ, 2021).

3.4 Conexões com o cenário internacional

Diversos países têm adotado legislações específicas para lidar com manipulação digital avançada. A União Europeia aprovou o Digital Services Act (DSA), que impõe obrigações mais rigorosas a grandes plataformas, incluindo mecanismos

para identificar e mitigar riscos sistêmicos como desinformação e manipulação coordenada (European Union, 2022).

Nos Estados Unidos, o Congresso tem buscado lidar com a manipulação digital por meio de propostas legislativas. Projetos como o *DEEPFAKES Accountability Act (U.S. Congress, 2019)*, introduzido no Congresso, visavam estabelecer requisitos de rotulagem obrigatória para conteúdos sintéticos. Embora o projeto não tenha sido aprovado, ele sinaliza a tendência regulatória federal em busca de transparência para o conteúdo gerado por IA.

3.5 Aspectos Relevantes

O arcabouço jurídico brasileiro oferece bases para a proteção contra práticas de manipulação digital, mas enfrenta limitações frente à sofisticação e velocidade das novas ameaças. A conjugação de normas como o Marco Civil, a LGPD e eventuais atualizações legislativas, somada à atuação proativa do Judiciário, é essencial para reduzir os impactos negativos à segurança da informação e aos direitos fundamentais.

Contudo, medidas exclusivamente repressivas não são suficientes; torna-se imprescindível investir em educação digital, transparência algorítmica e cooperação internacional para prevenir e conter a propagação dessas práticas (ENISA, 2023; Europol, 2024).

4 ESTUDO DE CASOS REAIS E ANÁLISE CRÍTICA

A análise de situações reais ajuda a compreender como as técnicas de manipulação digital deixam de ser conceitos abstratos e passam a afetar diretamente a política, o ambiente corporativo e as relações sociais. Ao observar como *deepfakes*, *bots* e redes coordenadas foram aplicados em contextos distintos, torna-se possível visualizar o alcance dessas tecnologias e os danos que podem causar. Cada caso evidencia, de forma concreta, como vulnerabilidades humanas e institucionais são exploradas pela engenharia social mediada por inteligência artificial, revelando a necessidade de respostas mais rápidas, eficientes e integradas.

4.1 Caso 1 – *Deepfake* em contexto político-eleitoral

Em setembro de 2023, durante o período pré-eleitoral na Eslováquia, circulou nas redes sociais um áudio supostamente atribuído a um candidato de oposição, no qual este discutia estratégias para manipular votos e subornar jornalistas. Posteriormente, verificou-se que o conteúdo fora gerado por inteligência artificial generativa, utilizando-se de *voice cloning* a partir de discursos públicos disponíveis na internet. O áudio foi disseminado por meio de grupos de mensageria e redes sociais, alcançando milhares de compartilhamentos antes de ser desmentido por veículos de imprensa e autoridades eleitorais (Techradar Pro, 2025).

A análise desse episódio demonstra como os *deepfakes* em formato de áudio — mais fáceis de produzir e mais difíceis de detectar visualmente — representam risco significativo para a integridade do processo democrático (Pawelec, 2022). Do ponto de vista da segurança da informação, houve violação do princípio da autenticidade, uma vez que a origem e a veracidade da mensagem foram falsificadas com alto grau de realismo (Shu *et al*., 2022). Sob a ótica jurídica, o caso evidencia a insuficiência de mecanismos normativos para retirada imediata de conteúdos altamente virais em períodos críticos, como o eleitoral, sem que haja risco de censura prévia (European Union, 2022).

4.2 Caso 2 – Golpe corporativo com deepfake em vídeo ao vivo

Em fevereiro de 2024, uma empresa multinacional com sede em Hong Kong reportou uma fraude de aproximadamente US$ 25 milhões, após um funcionário de finanças participar de uma reunião virtual via videoconferência na qual acreditava estar interagindo com o diretor financeiro e outros executivos da matriz. Na verdade, todos os participantes, exceto o funcionário, eram avatares gerados por deepfake em tempo real, com base em gravações públicas e internas obtidas por vazamentos (Cincodias, 2025).

O ataque, planejado com detalhamento, explorou vulnerabilidades humanas — confiança hierárquica e validação visual — e falhas procedimentais, como a ausência de confirmação por canais alternativos. Este episódio exemplifica como as novas formas de engenharia social potencializadas por IA generativa rompem com a lógica tradicional de autenticação, inclusive em ambientes corporativos que historicamente confiavam na validação visual ou na verificação vocal como mecanismos de segurança. Com a possibilidade de manipulação convincente dessas dimensões por deepfakes e modelos avançados de síntese de voz (Schmitt; Flechais, 2023), a fragilidade dos processos internos torna-se evidente. A análise de casos dessa natureza tem levado órgãos e especialistas a enfatizar a necessidade de protocolos de diligência reforçados, especialmente no que diz respeito à verificação de identidade, registro de eventos e preservação imediata de evidências digitais, conforme orienta o Conselho Nacional de Justiça em suas diretrizes de cadeia de custódia e prova digital (CNJ, 2021).

4.3 Caso 3 – Milícias digitais e desinformação coordenada no Brasil

No contexto brasileiro, investigações conduzidas pelo Supremo Tribunal Federal (Inquérito nº 4.781/DF) e pelo Tribunal Superior Eleitoral identificaram redes organizadas de perfis falsos e *bots* utilizados para disseminar desinformação, ataques pessoais e teorias conspiratórias (STF, 2020; TSE, 2021). As chamadas “milícias digitais” têm sido descritas pelo Tribunal Superior Eleitoral como estruturas articuladas que atuam de maneira coordenada para manipular o debate público, amplificar artificialmente determinadas narrativas e influenciar a opinião social por meio do uso estratégico de redes sociais e aplicativos de mensageria. Conforme documentado no Relatório de Enfrentamento à Desinformação nas Eleições de 2022, esses grupos

operam com organização interna, divisão de funções e mecanismos de impulsionamento capazes de potencializar conteúdos desinformativos, gerando efeitos significativos sobre a formação da vontade política do eleitorado (TSE, 2022).

Do ponto de vista da segurança da informação, essas operações afetam a integridade do ecossistema informacional, criando um ambiente onde informações falsas competem em igualdade — ou até com vantagem — frente a informações verificadas (Shu *et al*., 2022). Juridicamente, o caso impulsionou debates sobre a aplicação de dispositivos do Marco Civil da Internet e da legislação eleitoral, bem como sobre a necessidade de criminalizar condutas específicas relacionadas à manipulação digital coordenada.

4.4 Caso 4 – *Spam* social automatizado e fraude em *marketplace*

Relatórios da Europol (2024) e de empresas de cibersegurança apontam um aumento significativo no uso de *spam* social automatizado para promover páginas fraudulentas em plataformas de comércio eletrônico. Em um dos casos documentados, redes de *bots* criavam perfis falsos que interagiam de forma convincente com potenciais compradores, respondendo dúvidas, fornecendo avaliações falsas e compartilhando *links* para páginas externas fraudulentas (Europol, 2024).

De acordo com a ENISA (2023), campanhas de engenharia social automatizadas estão cada vez mais explorando a confiança gerada pelo formato de interação humana e pela aparência de aceitação social, e frequentemente induzem vítimas a efetuar pagamentos por produtos inexistentes, aproveitando-se da escala e personalização viabilizadas por IA. Já o relatório da Palo Alto Networks (2025) observa que tais ataques afetam de modo direto a confidencialidade e a integridade das relações comerciais digitais, com uso de respostas automatizadas cada vez mais refinadas e difíceis de distinguir de interações humanas.

4.5 Análise comparativa dos casos

A análise dos casos evidencia padrões comuns e particularidades:

- Padrões comuns: exploração de vulnerabilidades humanas (confiança, urgência, autoridade), uso de IA para escalar e sofisticar ataques, e dificuldade de detecção imediata (Shu *et al*., 2022).
- Particularidades:
  - Casos políticos demandam respostas rápidas para evitar impacto irreversível na opinião pública (Pawelec, 2022).
  - Casos corporativos envolvem altos valores e implicações contratuais e de governança (Techradar Pro, 2025).
  - As milícias digitais representam um desafio regulatório relevante justamente porque operam na zona cinzenta entre o discurso político legítimo e a desinformação deliberada. O *Programa Permanente de Enfrentamento à Desinformação – Eleições 2022* evidencia que essas estruturas exploram a velocidade e a opacidade das interações nas plataformas digitais para difundir mensagens manipuladas sob aparente roupagem de participação política, dificultando a atuação normativa e o controle institucional (TSE, 2022).
  - Fraudes comerciais afetam diretamente consumidores e exigem coordenação entre plataformas e órgãos de defesa do consumidor (Europol, 2024).

4.6 Aspectos Relevantes

Os exemplos apresentados demonstram que a engenharia social mediada por inteligência artificial constitui uma ameaça multifacetada, combinando elementos técnicos e manipulação psicológica em escala, com potencial de comprometer de forma profunda a segurança da informação (Shu et al., 2022; Schmitt; Flechais, 2023). A natureza transnacional e a velocidade de disseminação desses ataques dificultam a atuação exclusiva de autoridades nacionais, impondo a necessidade de cooperação internacional, integração legislativa e fortalecimento das capacidades de detecção (European Union, 2022; ENISA, 2023).

A partir desses casos, torna-se evidente que a resposta jurídica precisa ser acompanhada de estratégias preventivas, de conscientização e de adoção de boas práticas por usuários, empresas e instituições públicas.

5 MEDIDAS PREVENTIVAS E EDUCATIVAS

A análise de situações reais ajuda a compreender como as técnicas de manipulação digital deixam de ser conceitos abstratos e passam a afetar diretamente a política, o ambiente corporativo e as relações sociais. Ao observar como deepfakes, *bots* e redes coordenadas foram aplicados em contextos distintos, torna-se possível visualizar o alcance dessas tecnologias e os danos que podem causar. Cada caso evidencia, de forma concreta, como vulnerabilidades humanas e institucionais são exploradas pela engenharia social mediada por inteligência artificial, revelando a necessidade de respostas mais rápidas, eficientes e integradas.

5.1 Necessidade de abordagem multidimensional

A prevenção das novas formas de engenharia social mediadas por inteligência artificial requer atuação simultânea em três eixos: jurídico-regulatório, tecnológico e educacional (Shu *et al*., 2022). Tais eixos devem operar de forma coordenada, evitando tanto lacunas que possam ser exploradas por agentes maliciosos quanto sobreposições que gerem insegurança jurídica ou custos desnecessários para usuários e empresas (ENISA, 2023).

No plano jurídico, é importante consolidar e atualizar o arcabouço normativo, garantindo maior clareza quanto às responsabilidades de criadores, disseminadores e plataformas que hospedam conteúdos manipulados, inclusive no que se refere aos mecanismos de prevenção, mitigação e resposta a práticas de desinformação identificadas no ecossistema digital (TSE, 2022). No âmbito tecnológico, é imprescindível incentivar o desenvolvimento e a adoção de sistemas capazes de detectar *deepfakes, bots* e campanhas coordenadas com maior precisão, reconhecendo, porém, que tais ferramentas possuem limitações estruturais e dependem de transparência metodológica para preservar sua credibilidade (Schmitt; Flechais, 2023). Já no aspecto educacional, cabe investir em programas contínuos de alfabetização midiática (*media literacy*), capacitando cidadãos para identificar e reagir criticamente a conteúdos potencialmente manipulados (Pawelec, 2022).

5.2 Medidas jurídicas e regulatórias

Entre as medidas que podem ser implementadas no campo jurídico, destacam-se:

- Aprovação de legislação específica sobre conteúdos sintéticos, incluindo exigência de rotulagem (*watermarking*) de *deepfakes* e responsabilização pelo uso indevido de imagem ou voz (U.S. Congress, 2019).
- Regulamentação clara do uso de contas automatizadas em redes sociais, com obrigatoriedade de identificação como *bot* e rastreabilidade mínima para apuração de abusos (Brasil, 2020).
- Ampliação da cooperação entre órgãos nacionais (Polícia Federal, Ministério Público, ANPD, Justiça Eleitoral) e internacionais para investigação e remoção célere de conteúdos ilícitos transnacionais (Europol, 2024).
- Inclusão de diretrizes específicas para preservação de prova digital, permitindo atuação célere para coleta e guarda de evidências antes que sejam apagadas, modificadas ou ocultadas, conforme orienta o Conselho Nacional de Justiça em seu guia de boas práticas para produção e custódia da prova digital (CNJ, 2021).

5.3 Medidas tecnológicas

A tecnologia pode ser aliada estratégica na mitigação dos riscos, desde que adotada com respeito à privacidade e à liberdade de expressão. Algumas iniciativas incluem:

- Ferramentas de verificação de autenticidade de imagens, áudios e vídeos, utilizando técnicas de análise forense e detecção de manipulação (ENISA, 2023).
- Sistemas de monitoramento automatizado de comportamento de contas, capazes de identificar padrões típicos de *bots* ou de campanhas coordenadas (Palo Alto Networks, 2025).

- Adoção, por parte de plataformas, de alertas contextuais quando há suspeita de conteúdo manipulado, fornecendo ao usuário informações adicionais antes de compartilhar (Techradar Pro, 2025).
- Desenvolvimento de algoritmos de detecção abertos e auditáveis, evitando o risco de censura indevida e assegurando a confiança do público (Shu *et al*., 2022).

5.4 Medidas educacionais e de conscientização

O fortalecimento da resiliência social contra a manipulação digital depende, em grande medida, da capacidade crítica dos usuários. Nesse sentido, recomenda-se:

- Inclusão de conteúdos de alfabetização midiática nos currículos escolares, com abordagem prática sobre verificação de fontes e identificação de manipulação (Pawelec, 2022).
- Campanhas públicas de conscientização sobre riscos de *deepfakes*, milícias digitais e *spam* social, incentivando a checagem de informações antes do compartilhamento (ENISA, 2023).
- Parcerias entre governo, sociedade civil e empresas de tecnologia para disponibilizar guias e ferramentas de uso simples para validação de conteúdos (European Union, 2022).
- Treinamentos direcionados a jornalistas, comunicadores e figuras públicas constituem uma das linhas estratégicas previstas pelo Tribunal Superior Eleitoral no enfrentamento à desinformação, especialmente porque esses grupos figuram entre os alvos mais frequentes de campanhas de manipulação informacional. O Programa Permanente de Enfrentamento à Desinformação – Eleições 2022 destaca a importância de ações educativas específicas, voltadas à capacitação desses atores para identificar conteúdos enganosos, compreender técnicas de manipulação e adotar práticas seguras de comunicação digital (TSE, 2022).

5.5 Síntese das medidas

As medidas preventivas e educativas devem ser integradas, formando um ecossistema de proteção que atue tanto na origem (impedindo ou dificultando a produção e difusão de conteúdos ilícitos) quanto no destino (reduzindo a eficácia de tentativas de manipulação). Somente a conjugação desses esforços permitirá um ambiente informacional mais seguro e confiável, preservando direitos fundamentais e fortalecendo a democracia digital (Shu *et al*., 2022; ENISA, 2023; Europol, 2024).

6 CONSIDERAÇÕES FINAIS

A presente pesquisa analisou as novas formas de engenharia social mediadas por inteligência artificial em redes sociais, com destaque para *deepfakes*, *spam* social automatizado e milícias digitais. Verificou-se que tais práticas, embora distintas nas técnicas e nos objetivos imediatos, compartilham características que as tornam especialmente perigosas: elevada capacidade de persuasão, difícil detecção inicial e potencial de disseminação massiva em curto espaço de tempo (Shu *et al*., 2022).

O estudo demonstrou que o arcabouço jurídico brasileiro — composto pelo Marco Civil da Internet, pela Lei Geral de Proteção de Dados, por dispositivos do Código Penal e por iniciativas legislativas como o PL 2630/2020 — oferece bases relevantes para o enfrentamento das práticas de manipulação informacional. Contudo, conforme indicam análises institucionais do Tribunal Superior Eleitoral, esse conjunto normativo ainda demanda aperfeiçoamentos específicos para lidar com a sofisticação e a velocidade das ameaças emergentes no ambiente digital, especialmente aquelas potencializadas por sistemas de inteligência artificial (TSE, 2022). A jurisprudência, embora já apresente decisões relevantes, encontra desafios probatórios significativos, especialmente diante da volatilidade e do realismo crescente dos conteúdos manipulado. (TSE, 2019; TJSP, 2021).

A análise de casos reais evidenciou que os impactos dessas práticas ultrapassam a esfera individual, afetando empresas, instituições públicas e a própria integridade do processo democrático (Pawelec, 2022; Europol, 2024). Diante desse cenário, medidas preventivas e educativas, associadas a avanços regulatórios e tecnológicos, mostram-se essenciais para reduzir riscos e mitigar danos (Shu *et al*., 2022; ENISA, 2023).

O trabalho conclui que a proteção da segurança da informação, em seu sentido mais amplo, depende de um equilíbrio entre repressão e prevenção, com ações coordenadas entre Estado, setor privado e sociedade civil. A alfabetização midiática emerge como elemento-chave para formar cidadãos capazes de navegar com criticidade no ambiente digital, enquanto a regulação e a tecnologia devem garantir que tais cidadãos atuem em um ecossistema informacional minimamente íntegro e seguro (Pawelec, 2022; European Union, 2022).

Como agenda para futuras pesquisas, recomenda-se o aprofundamento de estudos sobre a eficácia de tecnologias de detecção de *deepfakes* em tempo real, a análise comparativa de legislações estrangeiras e o monitoramento contínuo da evolução das táticas de manipulação digital. Somente com atualização contínua do conhecimento técnico e das ferramentas de proteção será possível enfrentar os desafios resultantes da convergência entre inteligência artificial, redes sociais e segurança da informação, fenômeno que se intensifica com a sofisticação dos ataques mediatos por IA (Schmitt; Flechais, 2023; ENISA, 2023).

O avanço da inteligência artificial generativa indica que os próximos desafios já não se limitarão à detecção de fraudes digitais, mas à redefinição dos próprios fundamentos sobre os quais o direito constrói certezas. Em um ambiente informacional radicalmente manipulável, o conceito tradicional de prova tende a perder sua ancoragem objetiva, exigindo que o processo judicial abandone a confiança intuitiva na evidência audiovisual e adote novos parâmetros de autenticidade, verificabilidade e rastreabilidade algorítmica. A manipulação sintética deixa de ser apenas um problema técnico para se tornar um problema constitucional: a erosão da confiança pública — condição de possibilidade da democracia — ameaça o devido processo legal, a tutela jurisdicional efetiva e a própria legitimidade das instituições. Arriscamos um futuro em que a única verdade socialmente aceita será aquela produzida e transmitida por canais físicos ou por entidades dotadas de autoridade inquestionável, fragmentando ainda mais o tecido social e minando as bases da economia digital. A experiência internacional sugere que, em breve, os Estados terão de regular não apenas o uso de IA, mas o próprio direito à realidade, assegurando que cidadãos possam distinguir com segurança o que é autêntico do que é fabricado por sistemas autônomos. Assim, o futuro da segurança da informação não reside apenas em novas tecnologias de defesa, mas na capacidade de o ordenamento jurídico reconstruir critérios de verdade em um mundo em que verdade e falsidade se tornam produtos igualmente replicáveis e distribuíveis em escala global.

REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO/IEC 27000:2020**: Tecnologia da informação — Técnicas de segurança — Sistemas de gestão da segurança da informação — Visão geral e vocabulário. Rio de Janeiro: ABNT, 2020.

BIBRI, Simon Elias; HUANG, Jeffrey. ***Generative AI of things for sustainable smart cities: Synergizing cognitive augmentation, resource efficiency, network traffic, cybersecurity, and anomaly detection for environmental performance. Sustainable Cities and Society***, 2025.

BRASIL. Tribunal Superior Eleitoral. *Programa Permanente de Enfrentamento à Desinformação no âmbito da Justiça Eleitoral: Plano Estratégico – Eleições 2022*. Brasília: TSE, 2022. Disponível em: https://www.justicaeleitoral.jus.br/desinformacao/arquivos/programa-permanente-de-enfrentamento-a-desinformacao-novo.pdf. Acesso em: 14 nov. 2025.

BRASIL. Marco Civil da Internet. Lei nº 12.965, de 23 de abril de 2014. Brasília: Presidência da República, 2014. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Acesso em: 14 nov. 2025.

BRASIL. Lei Geral de Proteção de Dados Pessoais – *LGPD*. Lei nº 13.709, de 14 de agosto de 2018. Brasília: Presidência da República, 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 14 nov. 2025

BRASIL. Lei nº 14.155, de 27 de maio de 2021. Altera o Código Penal para prever crimes de fraude eletrônica. Diário Oficial da União, Brasília, DF, 28 maio 2021.

BRASIL. Supremo Tribunal Federal. Inquérito nº 4.781/DF. Rel. Min. Alexandre de Moraes. Brasília, DF, 2020.

BRASIL. Conselho Nacional de Justiça. **Cartilha de Boas Práticas para a Preservação e a Produção da Prova Digital**. Brasília: CNJ, 2021. Disponível em: https://www.cnj.jus.br/wp-content/uploads/2021/06/Cartilha-Prova-Digital-CNJ.pdf. Acesso em: 19 nov. 2025.

BRASIL. Conselho Nacional de Justiça**. Manual de Prova Digital para o Poder Judiciário. Brasília**: CNJ, 2021. Disponível em: https://www.cnj.jus.br. Acesso em: 19 nov. 2025.

BRASIL. Lei nº 14.155, de 27 de maio de 2021. Altera o Decreto-Lei nº 2.848, de 7 de dezembro de 1940 (Código Penal), e a Lei nº 7.210, de 11 de julho de 1984 (Lei de Execução Penal), para dispor sobre crimes cometidos mediante dispositivos eletrônicos. Disponível em: https://www.planalto.gov.br. Acesso em: 19 nov. 2025.

CINCODIAS. ***Spam en tiempo real: el nuevo arma de los ciberdelincuentes. El País***, 5 ago. 2025. Disponível em: https://cincodias.elpais.com. Acesso em: 19 ago. 2025.

ENISA – EUROPEAN UNION AGENCY FOR CYBERSECURITY. ***Threat Landscape 2023***. Disponível em: https://www.enisa.europa.eu. Acesso em: 19 ago. 2025.

ESTADOS UNIDOS. Congresso. ***H.R. 3230: DEEPFAKES Accountability Act of 2019***. *116th Congress. Washington, D.C.: U.S. Government Publishing Office, 2019*. Disponível em: https://www.congress.gov/bill/116th-congress/house-bill/3230. Acesso em: 20 nov. 2025.

EUROPEAN UNION. ***Digital Services Act: Ensuring transparency and accountability in online platforms. Brussels: European Commission***, 2022. Disponível em: https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package. Acesso em: 19 ago. 2025.

EUROPOL. ***Internet Organised Crime Threat Assessment (IOCTA)*** 2024. Disponível em: https://www.europol.europa.eu. Acesso em: 19 ago. 2025.

GOODFELLOW, ***Ian et al. Generative adversarial nets. In: Advances in Neural Information Processing Systems 27***. Montreal: Curran Associates, 2014. Disponível em: https://papers.neurips.cc/paper/5423-generative-adversarial-nets.pdf. Acesso em: 19 nov. 2025.

MITNICK, Kevin; SIMON, William. **A Arte de Enganar.** Tradução de Elizabeth L. de O. H. C. T. de la Santa. São Paulo: Pearson Education do Brasil, 2003.

PAWELEC, Maria. ***Deepfakes and democracy (theory): how synthetic audio-visual media for disinformation and hate speech threaten core democratic functions.*** Digital Society, v. 1, n. 19, p. 1-13, 2022. Disponível em: https://link.springer.com/article/10.1007/s44206-022-00010-6. Acesso em: 19 ago. 2025.

PALO ALTO NETWORKS – UNIT 42. ***Global Incident Response Report – Social Engineering Edition***. 2025. Disponível em: https://unit42.paloaltonetworks.com/2025-unit-42-global-incident-response-report-social-engineering-edition/. Acesso em: 19 ago. 2025.

RUSSELL, Stuart; NORVIG, Peter. **Inteligência Artificial: Uma Abordagem Moderna.** Tradução da Terceira Edição. São Paulo: Elsevier, 2013

SCHMITT, Marc; FLECHAIS, Ivan. ***Digital Deception: Generative Artificial Intelligence in Social Engineering and Phishing. arXiv preprint***, 2023. Disponível em: https://arxiv.org/abs/2310.13715. Acesso em:17 nov 2025

SHU, K. *et al*. ***The role of artificial intelligence in combating disinformation*. *Nature Human Behaviour*,** 2022. Disponível em: Disponível em: https://www.cambridge.org/core/journals/data-and-policy/article/role-of-artificial-intelligence-in-disinformation/7C4BF6CA35184F149143DE968FC4C3B6. Acesso em: 19 ago. 2025

STALLINGS, ***William. Cryptography and Network Security: Principles and Practice***. 7. ed. Boston: Pearson, 2017.

TECHRADAR PRO. ***Inside the deepfake threat that's reshaping corporate risk*.** Disponível em: https://www.techradar.com/pro. Acesso em: 19 ago. 2025.