www.revistaft.com.br

Revista FT | ISSN 1678-0817 | v. 30, n. 158, 2026 | DOI: 10.69849/a4syn865

# A inteligência artificial como ferramenta de manipulação e fraude: impactos sociais e desafios regulatórios na era digital

Artificial intelligence as a tool for manipulation and fraud: social impacts and regulatory challenges in the digital age

Caio Gabriel Correia da Silva Santos[1]; Bruna Silva Moreira[1]; Tiago José de Souza Cavalcanti [2]; João Pedro Pessoa e Silva Borba[1]; Luiz Eduardo Pereira dos Santos Júnior[1]; Samuel Silva Souza[1]; Adriano dos Santos Reis[3]; Gabriel Sena da Silva[4]

## Resumo

A consolidação da sociedade digital e o avanço da inteligência artificial generativa transformaram profundamente as dinâmicas sociais, econômicas e comunicacionais contemporâneas, ao mesmo tempo em que ampliaram os riscos associados às fraudes digitais e à manipulação informacional. Este estudo analisa os impactos sociais, jurídicos e tecnológicos decorrentes do uso da inteligência artificial em práticas fraudulentas, com ênfase em phishing, deepfakes, spoofing e engenharia social. A pesquisa caracteriza-se como qualitativa, exploratória e bibliográfica, fundamentada em revisão interdisciplinar envolvendo estudos de direito digital, cibersegurança, sociologia da tecnologia e inteligência artificial. Os resultados demonstram que a IA generativa potencializa a sofisticação dos ataques cibernéticos ao ampliar a capacidade de manipulação psicológica, falsificação identitária e disseminação de desinformação em larga

[1] Graduando em Sistemas de Informação - Centro Universitário UNIME - Lauro de Freitas, Bahia, Brasil
[2] Graduando em Psicologia - Centro Universitário UNIME - Lauro de Freitas, Bahia, Brasil
[3] Mestre em Engenharia Elétrica- Universidade Federal da Bahia - Salvador - BA. ORCID - https://orcid.org/0009-0008-5788-7462
[4] Bacharel em Arquitetura e Urbanismo - Centro Universitário UNIME - Itabuna, Bahia, Brasil

escala. Observa-se ainda que grupos socialmente vulneráveis, como idosos, mulheres e adolescentes, tornam-se alvos preferenciais dessas práticas, sofrendo impactos financeiros, emocionais e reputacionais significativos. Conclui-se que os desafios impostos pela IA ultrapassam a dimensão técnica, exigindo estratégias multidimensionais fundamentadas em letramento digital, fortalecimento regulatório, governança algorítmica e desenvolvimento ético das tecnologias emergentes.

**Palavras-Chave:** Inteligência Artificial Generativa; Fraudes Digitais; Deepfakes; Cibersegurança; Direitos Digitais.

## Abstract

The consolidation of digital society and the advancement of generative artificial intelligence have profoundly transformed contemporary social, economic, and communicational dynamics while simultaneously increasing the risks associated with digital fraud and informational manipulation. This study analyzes the social, legal, and technological impacts arising from the use of artificial intelligence in fraudulent practices, with emphasis on phishing, deepfakes, spoofing, and social engineering. The research is qualitative, exploratory, and bibliographic, based on an interdisciplinary review involving studies in digital law, cybersecurity, sociology of technology, and artificial intelligence. The results demonstrate that generative AI enhances the sophistication of cyberattacks by increasing the capacity for psychological manipulation, identity falsification, and large-scale dissemination of disinformation. It is also observed that socially vulnerable groups, such as older adults, women, and adolescents, become preferred targets of these practices, suffering significant financial, emotional, and reputational impacts. It is concluded that the challenges imposed by AI extend beyond the technical dimension, requiring multidimensional strategies grounded in digital literacy, regulatory strengthening, algorithmic governance, and ethical development of emerging technologies.

**Keywords:** Generative Artificial Intelligence; Digital Fraud; Deepfakes; Cybersecurity; Digital Rights.

# INTRODUÇÃO

A consolidação da sociedade digital, intensificada nas últimas décadas, tem produzido profundas transformações nas dinâmicas sociais, econômicas, culturais e políticas, ao mesmo tempo em que inaugura um campo complexo de tensões éticas, jurídicas e tecnológicas. Nesse contexto, a centralidade das Tecnologias da Informação e Comunicação (TICs) redefine não apenas os modos de interação humana, mas também as formas de acesso a direitos e serviços essenciais, configurando o que se convencionou denominar de direitos digitais, conceito ainda em construção e marcado por disputas teóricas, normativas e práticas. A expansão tecnológica, entretanto, não ocorre de maneira neutra ou homogênea, evidenciando a permanência das desigualdades digitais, manifestadas tanto na limitação de acesso à infraestrutura tecnológica quanto na ausência de competências para o uso qualificado dessas ferramentas. Tal cenário compromete a inserção social e produtiva de parcelas significativas da população, ampliando assimetrias historicamente presentes no capitalismo periférico. Nesse contexto, iniciativas locais de apropriação tecnológica, como as denominadas “tecnologias mundanas”, revelam estratégias de resistência e reinvenção desenvolvidas por comunidades marginalizadas, evidenciando que o uso da tecnologia constitui também um fenômeno socialmente situado (NEMER, 2021). A crescente utilização de algoritmos na mediação das relações sociais suscita questionamentos críticos acerca da suposta neutralidade das tecnologias digitais. Estudos demonstram que sistemas algorítmicos podem reproduzir e intensificar desigualdades estruturais, como racismo, exclusão social e discriminação de grupos minoritários, uma vez que operam a partir de bases de dados historicamente enviesadas (O’NEIL, 2020; NEMER, 2021). Dessa forma, a tecnologia deve ser compreendida como produto de contextos históricos, econômicos e culturais específicos, cujos impactos não podem ser dissociados das relações de poder que os atravessam. Outro aspecto central refere-se ao dilema contemporâneo entre segurança e privacidade no ambiente digital. Propostas regulatórias internacionais, como a Lei de Segurança Online do Reino Unido, evidenciam a elevada complexidade envolvida no equilíbrio entre a proteção contra crimes digitais – especialmente aqueles relacionados ao abuso infantil e às fraudes cibernéticas – e a garantia de direitos fundamentais, como privacidade, liberdade de expressão e sigilo das comunicações. Tal tensão demonstra a insuficiência de soluções simplistas e aponta para a necessidade de marcos regulatórios mais sofisticados, multidisciplinares e contextualmente sensíveis. Paralelamente, o avanço da Inteligência Artificial (IA), especialmente da IA

generativa, tem potencializado novas modalidades de criminalidade digital, destacando-se as deepfakes e as cripto fraudes. As deepfakes consistem em conteúdos sintéticos – vídeos, áudios ou imagens – manipulados por técnicas de inteligência artificial capazes de reproduzir rostos, vozes e comportamentos humanos com elevado grau de realismo. Já as cripto fraudes correspondem a esquemas criminosos que utilizam criptoativos, blockchain e serviços financeiros digitais para enganar usuários, desviar recursos e lavar dinheiro. Essas práticas exploram tanto a sofisticação tecnológica quanto a vulnerabilidade informacional das vítimas, ampliando significativamente os riscos de fraude, desinformação, manipulação social e instabilidade econômica (SUMSUB, 2024; NEGREIRO, 2025). Dados recentes demonstram crescimento expressivo dessas modalidades criminosas em escala global, com impactos financeiros significativos e aumento substancial das ocorrências no Brasil, impulsionados pela popularização de ferramentas de IA, pela digitalização acelerada das relações econômicas e pela instantaneidade dos meios de pagamento digitais. Nesse cenário, a fraude digital deixa de ser apenas um problema técnico e passa a configurar um fenômeno social complexo, atravessado por fatores econômicos, culturais, psicológicos e políticos. Diante desse contexto, torna-se imprescindível a construção de abordagens interdisciplinares capazes de articular direito, tecnologia e ciências sociais, visando ao desenvolvimento de uma ética digital apta a responder às ambivalências da contemporaneidade. A tecnologia, longe de ser neutra, constitui-se como espaço de disputa simbólica, econômica e material, exigindo não apenas regulação normativa, mas também formação crítica dos sujeitos, fortalecimento da educação digital e consolidação de estruturas democráticas no ambiente virtual.

## 1. O panorama das fraudes digitais e o impacto da IA generativa

A democratização de ferramentas de Inteligência Artificial generativa, aliada à instantaneidade de sistemas de pagamento como o Pix, alterou significativamente a dinâmica dos crimes cibernéticos no Brasil (SANTOS, 2025). Segundo Santos (2025), esse cenário tecnológico permite que agentes criminosos executem ataques cada vez mais sofisticados, personalizados e escaláveis, com menor esforço técnico e maior probabilidade de êxito. De forma simplificada, a fraude digital contemporânea ocorre na intersecção entre agentes motivados, vulnerabilidades estruturais dos sistemas de proteção e o uso de tecnologias baseadas em IA para personalização de abordagens, dificultando a identificação dos golpes por parte das vítimas. A utilização de

algoritmos capazes de simular voz, imagem e padrões comportamentais humanos amplia o potencial persuasivo das fraudes, tornando-as mais difíceis de detectar mesmo por usuários experientes. Entre as modalidades mais recorrentes, destacam-se aquelas fundamentadas na manipulação psicológica e na engenharia social para obtenção ilícita de dados pessoais, credenciais bancárias e transferências financeiras (SANTOS, 2025). Nesse contexto, técnicas como phishing, spoofing, clonagem de identidade digital e utilização de deepfakes vêm sendo amplamente empregadas para aumentar a credibilidade das abordagens criminosas e potencializar o alcance das fraudes em ambientes digitais. Atualmente, os golpes mais comuns são os seguintes:

## 1.1 Engenharia social, o direito e a tecnologia

A engenharia social constitui atualmente um dos principais mecanismos utilizados em fraudes digitais, explorando vulnerabilidades comportamentais e cognitivas dos usuários para obtenção ilícita de informações sensíveis, credenciais bancárias e acesso indevido a sistemas computacionais (BEZERRA; MILLIAN; SILVA, 2024). Diferentemente de ataques estritamente técnicos, essa modalidade criminosa fundamenta-se na manipulação psicológica das vítimas, utilizando estratégias de persuasão, urgência e confiança para induzir comportamentos favoráveis ao criminoso. Nesse contexto, a convergência entre tecnologia, direito e relações sociais torna-se elemento central para compreender a complexidade contemporânea dos crimes cibernéticos. 1.1.1 O crime por trás de um clique O termo phishing deriva do inglês fishing (“pescar”), metáfora que representa de maneira precisa a dinâmica desse tipo de fraude digital: o criminoso lança uma “isca”, geralmente na forma de e-mails, mensagens instantâneas ou páginas falsas, aguardando que a vítima “morda o anzol” ao clicar em links maliciosos ou fornecer informações pessoais (SOARES; RIBEIRO FILHO, 2022). Historicamente, práticas criminosas dependiam de proximidade física entre autor e vítima. Contudo, a expansão da internet e das plataformas digitais permitiu que criminosos atuassem de forma remota, escalável, relativamente anônima e com elevada rentabilidade econômica. O phishing destaca-se, nesse cenário, como uma modalidade baseada em engenharia social, compreendida como o conjunto de técnicas de manipulação psicológica utilizadas para induzir indivíduos a revelarem informações confidenciais ou executarem ações que comprometam sua segurança digital (SOARES; RIBEIRO FILHO, 2022). Além do roubo direto de credenciais bancárias e dados pessoais,

ataques dessa natureza frequentemente envolvem a instalação de malwares, o comprometimento de dispositivos e o acesso não autorizado a contas financeiras e redes corporativas. Dessa forma, o phishing ultrapassa a dimensão meramente tecnológica e passa a constituir um fenômeno social, psicológico e jurídico.

### 1.1.2 Anatomia e ciclo de vida do ataque

Os ataques de phishing não ocorrem de forma aleatória, mas seguem um ciclo estruturado e estrategicamente planejado, cujo objetivo é maximizar as chances de sucesso da fraude (BEZERRA; MILLIAN; SILVA, 2024). Em geral, esse processo envolve etapas interdependentes que articulam recursos tecnológicos e manipulação comportamental. A primeira etapa corresponde ao planejamento do ataque, momento em que o criminoso seleciona o alvo e desenvolve uma isca visualmente semelhante a instituições legítimas, como bancos, órgãos governamentais, plataformas de comércio eletrônico ou serviços digitais amplamente utilizados. A reprodução de elementos gráficos, logotipos e padrões de comunicação oficiais busca transmitir credibilidade e reduzir a desconfiança da vítima. Na sequência, ocorre a ativação do gatilho psicológico, considerada uma das etapas mais relevantes do processo fraudulento. O êxito do golpe depende diretamente da exploração de emoções humanas, especialmente medo, urgência, curiosidade ou ganância. Mensagens como “Sua conta será bloqueada”, “Atualize seus dados imediatamente” ou “Você recebeu um prêmio” são utilizadas para induzir respostas impulsivas e reduzir a capacidade crítica do usuário. A etapa final envolve a ação e coleta de dados. Nesse momento, a vítima clica em links maliciosos, acessa páginas clonadas ou realiza o download de arquivos comprometidos. Em páginas falsas, os dados inseridos são imediatamente capturados pelo criminoso. Já em arquivos infectados, pode ocorrer a instalação de programas maliciosos capazes de fornecer controle remoto do dispositivo, capturar senhas ou monitorar atividades digitais da vítima.

### 1.1.3 Modalidades:

Das técnicas clássicas às avançadas Com a evolução das tecnologias digitais, o phishing passou por um processo contínuo de sofisticação, adaptando-se a diferentes plataformas de comunicação e incorporando novas estratégias de ataque (SOARES; RIBEIRO FILHO, 2022). Entre as modalidades mais recorrentes, destacam-se os ataques direcionados, conhecidos como

spear phishing e whaling, nos quais criminosos selecionam vítimas específicas – frequentemente executivos, gestores ou funcionários estratégicos – utilizando informações previamente coletadas para aumentar a credibilidade da abordagem. Outras modalidades amplamente disseminadas são o smishing e o vishing. O primeiro ocorre por meio de mensagens SMS fraudulentas, enquanto o segundo utiliza chamadas telefônicas para simular centrais bancárias, órgãos públicos ou serviços de suporte técnico. Em ambos os casos, o objetivo é persuadir a vítima a compartilhar informações sigilosas ou realizar transferências financeiras. O pharming representa uma técnica ainda mais sofisticada, baseada na manipulação de servidores DNS. Nesse tipo de ataque, mesmo que o usuário digite corretamente o endereço eletrônico de uma instituição legítima, ele é redirecionado para páginas falsas controladas pelos criminosos. Já o tabnabbing explora o comportamento multitarefa dos usuários na navegação web: as abas inativas do navegador têm seu conteúdo alterado silenciosamente, exibindo posteriormente telas falsas de autenticação destinadas à captura de credenciais. A crescente sofisticação dessas modalidades demonstra que os ataques contemporâneos combinam vulnerabilidades técnicas e comportamentais, ampliando significativamente o potencial de dano das fraudes digitais.

### 1.1.4 O cenário brasileiro: dados e desafios legislativos

O Brasil figura entre os principais alvos mundiais de crimes cibernéticos. Dados recentes indicam que, apenas no primeiro semestre de 2020, foram registradas aproximadamente 2,6 bilhões de tentativas de ataques cibernéticos no país, correspondendo a cerca de 36% dos ataques identificados na América Latina (SANTOS, 2025). O processo acelerado de digitalização financeira, intensificado durante o período de isolamento social, associado à popularização de ferramentas como Pix, aplicativos bancários e sistemas digitais de cobrança, ampliou significativamente a superfície de ataque explorada por criminosos. Apesar do crescimento dessas práticas, o Direito Penal enfrenta dificuldades estruturais para acompanhar a velocidade das transformações tecnológicas. Segundo Souza, Rodrigues e Ferreira (2025), existe um relevante “abismo legislativo” entre a dinâmica dos crimes digitais e a capacidade normativa do Estado em regulá-los adequadamente. Embora instrumentos legais como a Lei Carolina Dieckmann (Lei nº 12.737/2012), o Marco Civil da Internet (Lei nº 12.965/2014) e a Lei Geral de Proteção de Dados – LGPD (Lei nº 13.709/2018) representem avanços importantes, ainda persistem lacunas relacionadas à responsabilização, rastreabilidade e prevenção de fraudes

digitais. Além das limitações normativas, fatores sociais também contribuem para a expansão dessas práticas criminosas. A vulnerabilidade econômica, o baixo nível de educação digital e o desconhecimento sobre mecanismos básicos de segurança favorecem a atuação de estelionatários e ampliam o número de vítimas em ambientes virtuais (MACEDO; SILVA, 2025).

### 1.1.5 Estratégias de defesa e prevenção

O enfrentamento das fraudes digitais exige estratégias de defesa em múltiplas camadas, articulando soluções tecnológicas, medidas jurídicas e ações educativas permanentes (BEZERRA; MILLIAN; SILVA, 2024). No âmbito técnico, destacam-se ferramentas como filtros de spam, protocolos de criptografia SSL/TLS, assinaturas digitais, autenticação em dois fatores (2FA) e sistemas de monitoramento comportamental baseados em inteligência artificial. Entretanto, embora os mecanismos tecnológicos sejam fundamentais, estudos apontam que a camada educacional permanece como o elemento mais decisivo na prevenção de ataques de engenharia social. Simulações práticas demonstram que muitos usuários continuam clicando em links suspeitos mesmo diante de alertas explícitos de segurança, evidenciando que o fator humano ainda constitui a principal vulnerabilidade explorada pelos criminosos. Nesse sentido, programas contínuos de conscientização em cibersegurança (security awareness) tornam-se essenciais para reduzir riscos e fortalecer a cultura de proteção digital. A adoção de práticas preventivas simples – como verificar remetentes, desconfiar de mensagens urgentes e evitar o compartilhamento de informações sensíveis em links não solicitados – constitui uma das formas mais eficazes de mitigação das fraudes digitais contemporâneas (BEZERRA; MILLIAN; SILVA, 2024).

## 2. A era da manipulação sintética: deepfakes, IA generativa e os novos desafios à verdade e ao direito

O ecossistema da desinformação passou, nas últimas décadas, por um processo de sofisticação tecnológica sem precedentes. O que anteriormente se manifestava predominantemente por meio de fake news – textos sensacionalistas, conteúdos descontextualizados e imagens manipuladas de maneira rudimentar – evoluiu para um cenário marcado pela utilização de mídias sintéticas altamente realistas (MOLINA; BERENGUEL,

2022). Nesse contexto, o avanço da Inteligência Artificial generativa permitiu a criação de conteúdos audiovisuais capazes de reproduzir expressões faciais, vozes e comportamentos humanos com elevado grau de precisão, dificultando a distinção entre realidade e manipulação. Grande parte dessa evolução decorre do desenvolvimento das Redes Generativas Adversariais (Generative Adversarial Networks – GANs), sistemas compostos por duas inteligências artificiais que operam de forma competitiva: enquanto uma rede é responsável por gerar conteúdos sintéticos, a outra atua identificando falhas e inconsistências, promovendo um processo contínuo de aperfeiçoamento (SOUSA et al., 2025). O resultado é a produção de vídeos, imagens e áudios que desafiam não apenas a percepção humana, mas também mecanismos tradicionais de autenticação digital. Na contemporaneidade, marcada pela lógica do "ver para crer", o impacto das imagens manipuladas tende a ser significativamente mais persuasivo do que textos falsos convencionais. A credibilidade atribuída ao conteúdo audiovisual potencializa processos de desinformação, manipulação política, fraudes financeiras e ataques à reputação, ampliando os riscos sociais associados ao uso indevido da IA generativa (MOLINA; BERENGUEL, 2022).

## 2.1 Formas de manipulação e engenharia social

A Inteligência Artificial generativa não apenas elevou a qualidade estética das fraudes digitais, mas também ampliou sua escala e capacidade de disseminação (SOUSA et al., 2025). O uso dessas tecnologias permitiu que ataques de engenharia social se tornassem mais personalizados, convincentes e difíceis de detectar, aumentando significativamente sua efetividade. Entre as modalidades emergentes, destacam-se os deepfakes em tempo real, já utilizados em videoconferências corporativas para simular executivos e autorizar operações financeiras fraudulentas. Casos recentes demonstram a utilização dessa técnica em fraudes milionárias, como o episódio registrado em Hong Kong, em 2024, no qual criminosos utilizaram avatares sintéticos para simular reuniões empresariais legítimas (SOUSA et al., 2025). A clonagem de voz e imagem representa outro fator crítico nesse processo. Ao reproduzir características biométricas de familiares, gestores ou representantes institucionais, a IA enfraquece mecanismos naturais de desconfiança das vítimas, favorecendo fraudes bancárias, transferências instantâneas via Pix e até burlas em sistemas de autenticação biométrica facial (SOUSA et al., 2025). Nesse cenário, a manipulação deixa de depender apenas da persuasão textual e passa a explorar vínculos emocionais e relações de confiança previamente estabelecidas.

Além das fraudes individuais, observa-se também o fortalecimento de estruturas coordenadas de manipulação informacional. Contas automatizadas e sistemas inteligentes de disseminação de conteúdo atuam na criação artificial de engajamento, propagação de narrativas falsas e ataques direcionados à reputação de indivíduos e instituições. Essas chamadas “milícias digitais” utilizam bots e algoritmos para ampliar artificialmente a circulação de conteúdos manipulados, influenciando debates públicos e potencializando campanhas de desinformação em larga escala (LUCAS, 2025).

## 2.2 Deepfakes e a violação dos direitos da personalidade

A utilização de deepfakes em contextos políticos e eleitorais representa uma das principais preocupações contemporâneas relacionadas à integridade democrática e à proteção da informação. Durante períodos eleitorais, conteúdos manipulados têm sido utilizados para criar “escândalos artificiais”, distorcer declarações públicas e produzir cenários fictícios destinados a influenciar emocionalmente o eleitorado e gerar instabilidade informacional (ANASTACIO; TAVARES, 2024). Embora práticas tradicionais de desinformação ainda predominem em muitos contextos, a democratização do acesso às ferramentas de IA generativa amplia significativamente o potencial de disseminação dessas estratégias. A facilidade de criação de vídeos e áudios sintéticos reduz custos operacionais, acelera a produção de conteúdo manipulado e intensifica o risco de campanhas coordenadas de desinformação política (ANASTACIO; TAVARES, 2024). Entretanto, os impactos dos deepfakes ultrapassam a esfera política e atingem diretamente os direitos da personalidade. A utilização indevida da imagem, da voz e da identidade de indivíduos compromete valores fundamentais relacionados à dignidade humana, à honra, à privacidade e à autodeterminação informacional. Segundo Martins (2024), a tecnologia deepfake retira do indivíduo o controle sobre sua própria representação digital, sendo frequentemente empregada em práticas de pornografia não consensual, difamação e atribuição fraudulenta de falas ou comportamentos inexistentes. Os danos decorrentes dessas práticas são frequentemente irreversíveis, sobretudo diante da velocidade de disseminação dos conteúdos em redes sociais e plataformas digitais. Mesmo quando removidos posteriormente, os materiais manipulados podem permanecer armazenados, compartilhados ou reproduzidos em diferentes ambientes virtuais, dificultando a reparação integral dos prejuízos morais e sociais causados às vítimas.

## 2.3 O vácuo legislativo e os desafios da prova

O ordenamento jurídico brasileiro ainda enfrenta dificuldades significativas para acompanhar a velocidade de evolução das tecnologias baseadas em Inteligência Artificial. Embora instrumentos normativos como o Marco Civil da Internet, a Lei Geral de Proteção de Dados (LGPD) e dispositivos do Código Penal representem avanços importantes no enfrentamento de crimes digitais, persistem lacunas relacionadas à responsabilização, rastreamento e produção de provas em casos envolvendo mídias sintéticas (MARTINS, 2024). Um dos principais desafios refere-se à identificação e responsabilização dos autores das fraudes digitais. O anonimato proporcionado por plataformas digitais, aliado à utilização de redes privadas, servidores internacionais e ferramentas de mascaramento de identidade, dificulta significativamente a localização da origem dos ataques e a atribuição de responsabilidade criminal (SOUSA et al., 2025). Além disso, emerge um problema jurídico ainda mais complexo: o dilema da prova digital. Em um cenário no qual vídeos, áudios e imagens podem ser produzidos artificialmente com elevado grau de realismo, a confiabilidade das evidências audiovisuais passa a ser progressivamente questionada. Conteúdos que historicamente eram considerados provas robustas em processos judiciais podem deixar de possuir presunção automática de autenticidade, exigindo novos métodos periciais de validação (LUCAS, 2025). Apesar dessas dificuldades, iniciativas institucionais já demonstram movimentos de adaptação do sistema jurídico brasileiro. Investigações como o Inquérito das Milícias Digitais, conduzido no âmbito do Supremo Tribunal Federal (STF), evidenciam a crescente preocupação com ataques coordenados de desinformação e manipulação digital, bem como a necessidade de fortalecimento de mecanismos jurídicos e tecnológicos capazes de enfrentar essas ameaças contemporâneas.

## 2.4 Estratégias de defesa e segurança pública

O enfrentamento das ameaças associadas aos deepfakes e à IA generativa exige uma abordagem multidimensional, envolvendo integração entre tecnologia, direito, políticas públicas e educação digital (SOUSA et al., 2025). Não se trata apenas de um problema técnico, mas de um desafio estrutural relacionado à proteção da confiança social e à preservação da integridade informacional no ambiente digital. No campo da segurança pública e da investigação criminal, destaca-se a necessidade de fortalecimento da perícia forense digital. O desenvolvimento de

ferramentas automatizadas capazes de identificar inconsistências imperceptíveis ao olhar humano – como padrões anômalos de iluminação, sincronização labial, compressão de imagem e frequência vocal – torna-se essencial para detectar conteúdos manipulados (MOLINA; BERENGUEL, 2022). Nesse contexto, surgem sistemas de IA especializados na identificação de deepfakes, funcionando como mecanismos de verificação técnica e autenticação digital. Outro aspecto relevante refere-se à transparência algorítmica nas plataformas digitais. Especialistas defendem a implementação de mecanismos obrigatórios de identificação e rotulagem de conteúdos sintéticos, permitindo que usuários reconheçam materiais produzidos ou alterados por Inteligência Artificial (SOUSA et al., 2025). Medidas dessa natureza podem contribuir para reduzir a circulação de conteúdos fraudulentos e ampliar a responsabilização das plataformas digitais. Entretanto, diante da velocidade de evolução tecnológica, o letramento digital permanece como uma das principais estratégias de defesa social. Segundo Lucas (2025), a formação crítica dos usuários constitui a última linha de proteção contra a manipulação informacional. Nesse sentido, a educação midiática tornase indispensável para desenvolver competências relacionadas à verificação de fontes, análise contextual de conteúdos e identificação de sinais de desinformação, independentemente do grau de realismo apresentado pelos materiais audiovisuais. Portanto, a construção de ambientes digitais mais seguros depende não apenas de avanços tecnológicos e regulatórios, mas também da consolidação de uma cultura de cidadania digital crítica, consciente e preparada para enfrentar os desafios impostos pela era da manipulação sintética.

## 3. A máscara da fraude: spoofing, engenharia social e os desafios da segurança cibernética

O cenário contemporâneo das fraudes digitais possui suas bases estruturais nas vulnerabilidades históricas das redes de comunicação. Entre as principais técnicas utilizadas nesse contexto destaca-se o spoofing, mecanismo que permite a falsificação de informações de identificação digital, como endereços IP, e-mails, domínios e números telefônicos (Caller ID) (NASCIMENTO, 2020). Por meio dessa técnica, criminosos conseguem mascarar sua verdadeira identidade digital e simular comunicações legítimas, ampliando significativamente a eficácia de ataques de engenharia social. A fragilidade desse sistema decorre, em grande parte, da própria arquitetura histórica das redes de telecomunicações. Protocolos antigos de sinalização, como o

SS7 (Signaling System No. 7), foram desenvolvidos priorizando funcionalidade, interoperabilidade e velocidade de comunicação, sem incorporar mecanismos robustos de autenticação e verificação de origem (NASCIMENTO, 2020). Em razão dessa lógica estrutural baseada na "confiança" entre sistemas, torna-se possível manipular informações de identificação com relativa facilidade. Com a popularização de tecnologias VoIP (Voice over Internet Protocol) e a disponibilidade de gateways de baixo custo, indivíduos mal-intencionados passaram a utilizar recursos capazes de simular números oficiais de bancos, operadoras telefônicas e órgãos públicos. Dessa forma, o usuário recebe chamadas aparentemente legítimas, o que reduz sua capacidade de suspeita e aumenta o potencial de sucesso das fraudes (NASCIMENTO, 2020). Nesse contexto, o spoofing não representa apenas uma vulnerabilidade técnica isolada, mas um elemento estruturante das fraudes digitais contemporâneas, funcionando como mecanismo de legitimação artificial das abordagens criminosas.

## 3.1 O ecossistema de golpes no Brasil e a engenharia social

No Brasil, a combinação entre hiperconectividade, ampla utilização de redes sociais e popularização de sistemas de pagamento instantâneo, como o Pix, contribuiu significativamente para a expansão das fraudes digitais (SANTOS, 2025). A transformação digital acelerada das relações sociais e econômicas ampliou a superfície de ataque disponível aos criminosos, permitindo que golpes sejam aplicados em larga escala e com elevado potencial de monetização. Nesse cenário, observa-se uma mudança estratégica relevante: ao invés de concentrar esforços exclusivamente em ataques sofisticados contra servidores e sistemas corporativos, os criminosos passaram a direcionar suas ações diretamente ao usuário final. A engenharia social tornou-se, assim, o principal instrumento operacional dessas fraudes, explorando emoções humanas como medo, urgência, confiança e oportunidade para contornar barreiras tecnológicas de segurança (SANTOS, 2025). Entre os golpes mais recorrentes destacam-se a clonagem de contas de WhatsApp, fraudes envolvendo boletos bancários adulterados e diferentes modalidades de golpes via Pix, incluindo o chamado golpe da "Mão Fantasma", no qual a vítima é induzida a instalar softwares de acesso remoto em seu dispositivo (SANTOS, 2025). Essas práticas frequentemente utilizam dados pessoais previamente obtidos em vazamentos de informações, fortalecendo a credibilidade das abordagens criminosas. Os dados pessoais passaram a ocupar posição central nesse ecossistema ilícito, sendo frequentemente descritos como o "novo petróleo" da economia

digital. Informações cadastrais, registros bancários, credenciais de acesso e dados biométricos são comercializados ilegalmente em ambientes da Dark Web, alimentando um mercado clandestino altamente lucrativo e sofisticado (BEZERRA NETO; SIQUEIRA, 2025). Esse cenário evidencia que a fraude digital contemporânea está inserida em uma cadeia econômica transnacional baseada na coleta, circulação e exploração de dados sensíveis.

## 3.2 Alvos estratégicos e invasão de privacidade

O spoofing telefônico não se limita à aplicação de fraudes financeiras imediatas, funcionando também como porta de entrada para ataques mais complexos relacionados à invasão de dispositivos e comprometimento de contas digitais. Um exemplo emblemático envolve a exploração de sistemas de correio de voz associados a aplicativos de mensagens instantâneas. Nesse tipo de ataque, o criminoso utiliza técnicas de falsificação de identidade telefônica para interceptar códigos de autenticação enviados por serviços como Telegram e WhatsApp, assumindo o controle da conta da vítima (SPÍNOLA, 2020). A sofisticação desses ataques evidencia como mecanismos tradicionais de autenticação ainda apresentam fragilidades significativas diante da evolução das técnicas de engenharia social. Além da dimensão tecnológica, observa-se a existência de grupos particularmente vulneráveis a esse tipo de fraude. Entre os principais alvos estão pessoas idosas, que frequentemente apresentam menor familiaridade com ferramentas digitais e maior confiança em contatos telefônicos institucionais (MORGADO et al., 2023). Criminosos exploram essa vulnerabilidade combinando manipulação emocional, simulação de autoridade e urgência psicológica para induzir vítimas a realizar transferências financeiras, compartilhar códigos de autenticação ou fornecer informações sigilosas. Os impactos dessas fraudes ultrapassam os prejuízos financeiros imediatos. Muitas vítimas desenvolvem consequências psicológicas relevantes, incluindo ansiedade, medo, vergonha social e perda de confiança em sistemas digitais e instituições financeiras (MORGADO et al., 2023). Dessa forma, as fraudes digitais devem ser compreendidas também como fenômenos de violência simbólica e psicológica, capazes de comprometer a autonomia e a segurança subjetiva dos indivíduos.

## 3.3 O cenário jurídico e os desafios da investigação

A legislação brasileira ainda enfrenta dificuldades significativas para acompanhar a velocidade de transformação das ameaças cibernéticas contemporâneas. Dispositivos legais como o Art. 154-A do Código Penal, relacionado à invasão de dispositivos informáticos, e o Art. 307, referente ao crime de falsa identidade, representam tentativas de adaptação normativa ao contexto digital, mas frequentemente mostram-se insuficientes diante da sofisticação dos crimes praticados em ambientes virtuais (SPÍNOLA, 2020). Um dos principais desafios jurídicos refere-se à delimitação da proteção de dados armazenados em ambientes digitais distribuídos, especialmente serviços em nuvem. A natureza descentralizada dessas informações levanta questionamentos sobre jurisdição, competência investigativa e extensão das garantias legais relacionadas à privacidade e à inviolabilidade de dados. Além disso, o caráter transnacional do cibercrime dificulta significativamente os processos de investigação e responsabilização. Criminosos utilizam redes privadas virtuais (VPNs), servidores estrangeiros, criptomoedas e sistemas de anonimização para ocultar rastros digitais e reduzir possibilidades de rastreamento pelas autoridades competentes (BEZERRA NETO; SIQUEIRA, 2025). Outro fator crítico refere-se à insuficiência estrutural dos órgãos de segurança pública no enfrentamento dessas ameaças. A escassez de profissionais especializados, ferramentas forenses atualizadas e investimentos em inteligência cibernética amplia as dificuldades investigativas e reduz a capacidade estatal de resposta aos crimes digitais. Nesse contexto, a proteção de dados pessoais deixa de constituir apenas uma obrigação regulatória prevista na LGPD e passa a ser compreendida como questão estratégica de segurança pública e segurança nacional.

## 3.4 Mitigação e o futuro da defesa: IA e educação digital

O enfrentamento das fraudes baseadas em spoofing exige uma abordagem multidimensional, integrando soluções tecnológicas, medidas regulatórias e ações permanentes de conscientização social (NASCIMENTO, 2020). Em razão da sofisticação crescente dos ataques, estratégias isoladas tornam-se insuficientes para garantir proteção efetiva em ambientes digitais. No âmbito técnico, destacam-se iniciativas voltadas à implementação de protocolos avançados de autenticação de chamadas, como o STIR/SHAKEN, sistema que utiliza assinaturas digitais para validar a origem de comunicações telefônicas e reduzir fraudes envolvendo falsificação de números (NASCIMENTO, 2020). Paralelamente, discute-se a necessidade de maior rigor regulatório por parte da Agência Nacional de Telecomunicações (ANATEL), especialmente no

monitoramento de operadoras e serviços de telefonia digital. A Inteligência Artificial também emerge como importante ferramenta defensiva. Sistemas baseados em machine learning são capazes de identificar padrões anômalos de comportamento, detectar tentativas de fraude em tempo real e bloquear transações suspeitas em poucos milissegundos (MORGADO et al., 2023). Esses mecanismos permitem o cruzamento automatizado de dados comportamentais, geográficos e transacionais, aumentando significativamente a capacidade preventiva das instituições financeiras e plataformas digitais. Entretanto, apesar dos avanços tecnológicos, especialistas apontam que a educação digital continua sendo a principal barreira contra ataques de engenharia social. Programas tradicionais de conscientização vêm sendo substituídos por metodologias mais interativas, incluindo gamificação, simulações práticas e aplicativos educacionais capazes de reproduzir cenários reais de fraude (POLIDO, 2023). A construção de uma "desconfiança saudável" diante de contatos digitais, associada ao uso de mecanismos de autenticação em duas etapas (2FA), constitui uma das estratégias mais eficazes para redução de vulnerabilidades humanas exploradas pelos criminosos (SANTOS, 2025). Assim, o fortalecimento do capital humano e do letramento digital torna-se elemento indispensável para a consolidação de uma cultura de segurança cibernética mais resiliente e adaptada aos desafios contemporâneos.

## 4. Principais consequências dos golpes digitais para a sociedade

Após a análise das principais modalidades de golpes virtuais praticados atualmente, torna-se fundamental compreender as consequências sociais produzidas por essas fraudes digitais. Os crimes cibernéticos envolvendo Inteligência Artificial vêm se tornando cada vez mais frequentes e sofisticados, produzindo impactos que ultrapassam as perdas financeiras imediatas. Seus efeitos atingem dimensões relacionadas à segurança coletiva, à confiança social, à estabilidade institucional e ao bem-estar psicológico da população. A utilização de ferramentas de IA generativa, deepfakes, engenharia social e manipulação de dados potencializa a capacidade de disseminação desses golpes, ampliando sua abrangência e dificultando mecanismos tradicionais de identificação e proteção. Nesse contexto, observam-se consequências significativas que afetam não apenas indivíduos isolados, mas toda a estrutura social contemporânea.

### 4.1 Quebra de confiança e crise da informação

O termo deepfake refere-se a conteúdos audiovisuais artificiais – como vídeos, imagens e áudios – produzidos ou manipulados por Inteligência Artificial, simulando pessoas reais com elevado grau de realismo (PAYNE, 2024). A popularização dessas tecnologias contribuiu para a intensificação da disseminação de informações falsas e conteúdos adulterados, gerando uma crescente desconfiança social em relação às informações consumidas em ambientes digitais. A ampla circulação de deepfakes e notícias falsas inaugura uma crise de confiança informacional, na qual os indivíduos passam a questionar a autenticidade dos conteúdos que acessam diariamente (CASTRO; ZORZAN, 2026). Em uma sociedade fortemente baseada em mídias digitais, essa instabilidade compromete não apenas a circulação da informação, mas também relações sociais, institucionais e políticas. Como consequência, desenvolve-se um ambiente permanente de insegurança comunicacional, no qual usuários sentem necessidade constante de verificar a autenticidade de vídeos, imagens, mensagens e notícias. O uso das redes sociais, originalmente associado ao entretenimento, à interação social e ao aprendizado, passa a ser atravessado pelo medo de manipulação e fraude. Os conteúdos manipulados por IA apresentam níveis de realismo cada vez mais sofisticados, capazes de enganar inclusive usuários atentos e experientes (MONTEIRO, 2025). Nesse contexto, o phishing destaca-se como um dos golpes mais representativos dessa crise de confiança. Essa prática fraudulenta consiste no envio de mensagens que simulam comunicações oficiais de instituições legítimas – como bancos, lojas virtuais ou órgãos governamentais – induzindo usuários a compartilhar dados pessoais e financeiros de maneira involuntária (COMITÊ GESTOR DA INTERNET NO BRASIL, 2012). A recorrência dessas práticas fragiliza a confiança coletiva nas plataformas digitais e compromete a credibilidade das próprias instituições legítimas, que passam a enfrentar dificuldades adicionais para estabelecer comunicação segura com a população.

## 4.2 Danos à reputação e violação da imagem

Os golpes digitais baseados em Inteligência Artificial frequentemente utilizam imagens, vídeos e vozes de pessoas reais sem autorização, produzindo graves impactos à reputação e à honra das vítimas. A manipulação de conteúdos permite criar cenários falsos nos quais indivíduos aparentam dizer ou realizar ações que jamais ocorreram, promovendo intensa distorção da realidade. Segundo Molina e Berenguel (2022), os deepfakes possuem elevado potencial de comprometer reputações pessoais e profissionais, sobretudo quando utilizados com finalidade

ilícita, ofensiva ou difamatória. A facilidade de disseminação desses conteúdos em redes sociais amplia exponencialmente os danos causados às vítimas, dificultando mecanismos de reparação posterior. Nesse cenário, o uso indevido da Inteligência Artificial intensifica violações aos direitos da personalidade, especialmente ao direito de imagem, à honra e à privacidade. Gomes, Silva e Domingos (2025) destacam que as novas tecnologias de manipulação digital criam formas inéditas de violação de direitos fundamentais, ampliando os desafios jurídicos relacionados à responsabilização civil e criminal. Os impactos tornam-se ainda mais sensíveis em contextos políticos e eleitorais. Em ambientes marcados por elevada competitividade e polarização, conteúdos manipulados podem ser utilizados para desinformação, ataques à reputação de candidatos e manipulação da opinião pública. Batista e Santaella (2024) afirmam que a disseminação de notícias falsas e conteúdos adulterados compromete o funcionamento democrático e intensifica processos de polarização política. Azevedo (2025) ressalta que o avanço da Inteligência Artificial representa um novo desafio para campanhas eleitorais, permitindo a criação de conteúdos sintéticos que simulam discursos, ações e posicionamentos de candidatos. Nesse sentido, os deepfakes configuram uma nova forma de violência política digital, marcada pela manipulação da percepção pública e pela instrumentalização da desinformação como ferramenta de disputa política. Dessa forma, observa-se que os impactos reputacionais provocados pelas fraudes digitais ultrapassam a esfera individual, afetando também instituições, processos democráticos e relações de confiança social.

## 4.3 Danos psicológicos e emocionais

Além das consequências financeiras e reputacionais, os golpes digitais envolvendo Inteligência Artificial produzem impactos psicológicos e emocionais significativos nas vítimas. A exposição pública indevida, o sentimento de violação da intimidade e o medo constante de novas fraudes contribuem para o desenvolvimento de sofrimento psíquico intenso. Entre os crimes digitais mais nocivos nesse contexto destacam-se os chamados deep nudes, prática na qual imagens íntimas falsas são produzidas por meio de Inteligência Artificial utilizando fotografias extraídas de redes sociais pessoais (AGÊNCIA PÚBLICA, 2025). Essa modalidade de violência digital atinge principalmente mulheres e adolescentes, produzindo consequências severas relacionadas à saúde mental e ao convívio social. Relatos apresentados pela Agência Pública (2025) demonstram que vítimas desse tipo de fraude frequentemente descrevem sentimentos de

abuso, humilhação e perda de controle sobre a própria imagem. Alguns afirmam ter desenvolvido estados prolongados de paranoia, medo social e insegurança emocional após a divulgação dos conteúdos manipulados. Além disso, pesquisas acadêmicas apontam impactos ainda mais profundos. A tese de doutorado da socióloga Laís Patrocínio, da Universidade Estadual de Minas Gerais (UEMG), identificou consequências como depressão, transtornos alimentares, automutilação, isolamento social, ansiedade intensa e ideação suicida em vítimas de deepfakes e crimes digitais relacionados (AGÊNCIA PÚBLICA, 2025). Esses impactos não se restringem apenas aos deep nudes. Fraudes financeiras, roubo de identidade, cyberbullying, clonagem de voz e golpes baseados em engenharia social também geram sofrimento psicológico relevante, afetando a sensação de segurança e confiança dos indivíduos em ambientes digitais. Diante desse cenário, observa-se que a expansão dos crimes cibernéticos baseados em Inteligência Artificial contribui para o aumento da vulnerabilidade social e da insegurança coletiva. O rápido avanço tecnológico, aliado à dificuldade de adaptação dos mecanismos jurídicos e de proteção digital, intensifica a percepção de fragilidade da sociedade diante das novas formas de manipulação e violência informacional.

## 5. Públicos vulneráveis e o impacto da Inteligência Artificial

A chegada da Inteligência Artificial generativa representa não apenas um avanço tecnológico, mas uma transformação estrutural nas dinâmicas de segurança digital contemporâneas. Se anteriormente golpes virtuais podiam ser identificados por sinais relativamente perceptíveis – como erros gramaticais, links suspeitos ou comunicações visualmente precárias –, atualmente a IA permite a automatização de ataques altamente sofisticados, personalizados e adaptados ao perfil específico de cada vítima. Nesse contexto, a tecnologia deixa de ser apenas uma ferramenta operacional e passa a constituir um ambiente que reorganiza relações sociais, formas de comunicação e percepções de realidade. Conforme argumenta Pierre Lévy (1999), o ciberespaço não deve ser compreendido apenas como infraestrutura técnica, mas como espaço cultural capaz de redefinir práticas sociais, cognitivas e comunicacionais. A Inteligência Artificial amplia esse processo ao tornar cada vez mais difusa a distinção entre conteúdos autênticos e produções sintéticas. Essa transformação também pode ser compreendida à luz da noção de “modernidade líquida” proposta por Bauman (2001), marcada pela instabilidade das relações, pela fragilidade das referências tradicionais e pela fluidez das

identidades contemporâneas. No ambiente digital mediado por IA, a própria ideia de verdade passa a assumir caráter instável, vulnerável à manipulação algorítmica e à produção massiva de conteúdos artificiais. Dessa forma, a vulnerabilidade digital contemporânea não se limita à ausência de habilidades técnicas ou ao desconhecimento tecnológico. O indivíduo vulnerável é aquele inserido em um ambiente no qual mecanismos de manipulação se tornam suficientemente sofisticados para desafiar sentidos, emoções e capacidades críticas. A Inteligência Artificial amplia a assimetria entre usuários comuns e estruturas criminosas altamente organizadas, tornando a proteção digital um desafio cada vez mais complexo.

## 5.1 Idosos: a armadilha da voz e o afeto manipulado

Entre os grupos mais vulneráveis aos golpes digitais mediados por IA destacam-se os idosos, frequentemente alvo de fraudes que exploram vínculos afetivos, confiança interpessoal e fragilidades emocionais. A evolução dos golpes telefônicos tradicionais para modalidades sofisticadas de vishing com clonagem de voz evidencia o potencial destrutivo dessas tecnologias. A partir de pequenos fragmentos de áudio obtidos em redes sociais, aplicativos de mensagens ou vídeos públicos, criminosos conseguem reproduzir com elevado grau de realismo a voz de familiares próximos, como filhos ou netos. Utilizando mensagens de urgência emocional – pedidos de ajuda financeira, acidentes ou situações emergenciais –, os golpistas exploram relações de confiança para induzir transferências bancárias instantâneas via Pix. Dados da FEBRABAN (2026) indicam crescimento significativo da eficácia dessas fraudes contra pessoas com mais de 60 anos, especialmente em golpes que utilizam recursos de clonagem de voz baseados em IA. O problema central, nesse contexto, não reside apenas na sofisticação tecnológica do ataque, mas na instrumentalização do afeto e das relações familiares como mecanismos de manipulação. Diante desse cenário, estratégias preventivas simples assumem importância significativa. A criação de palavras-chave familiares para situações emergenciais, a confirmação de informações por múltiplos canais e o fortalecimento do apoio familiar tornam-se mecanismos fundamentais de proteção. Além disso, políticas públicas de inclusão e educação digital voltadas à população idosa revelam-se indispensáveis para redução dessas vulnerabilidades.

## 5.2 Crianças e adolescentes: o perigo por trás do algoritmo

Crianças e adolescentes também figuram entre os grupos mais expostos aos impactos da Inteligência Artificial no ambiente digital. Diferentemente das gerações anteriores, esses indivíduos nasceram inseridos em ecossistemas digitais altamente mediados por algoritmos, nos quais plataformas decidem continuamente quais conteúdos serão consumidos, priorizados e recomendados. Nesse contexto, os riscos tornam-se frequentemente invisíveis e naturalizados. Algoritmos de recomendação podem criar "bolhas informacionais" que influenciam comportamentos, percepções de mundo, construção de identidade e padrões de autoestima. A exposição contínua a conteúdos manipulados, hiper estimulantes ou violentos pode gerar impactos significativos no desenvolvimento emocional e cognitivo desse público. Entre as ameaças mais graves destaca-se a utilização de deepfakes envolvendo menores de idade. A SaferNet Brasil alerta para o crescimento alarmante da produção de montagens pornográficas geradas por Inteligência Artificial utilizando imagens de crianças e adolescentes extraídas de redes sociais. Ferramentas conhecidas como Deep Nude permitem a criação artificial de imagens íntimas falsas, frequentemente utilizadas para humilhação, cyberbullying, extorsão e violência sexual digital. Os impactos psicológicos dessas práticas podem ser duradouros, afetando autoestima, saúde mental, relações sociais e desenvolvimento emocional das vítimas. Além disso, a velocidade de circulação desses conteúdos dificulta processos de remoção e reparação dos danos causados. Diante dessa realidade, a educação digital contemporânea não pode restringir se ao ensino operacional das tecnologias. Torna-se necessário desenvolver competências críticas relacionadas à verificação de informações, compreensão do funcionamento algorítmico das plataformas e identificação de manipulações digitais. O letramento midiático e informacional passa a constituir elemento central da proteção de crianças e adolescentes em ambientes digitais.

## 5.3 Mulheres: a IA como ferramenta de violência de gênero

A Inteligência Artificial também intensifica formas históricas de violência de gênero já existentes no espaço social, ampliando mecanismos de assédio, exposição e ataque à reputação das mulheres no ambiente digital. Nesse contexto, a IA generativa funciona como ferramenta de potencialização de práticas discriminatórias e abusivas. Entre os golpes mais recorrentes destacam-se as fraudes românticas mediadas por modelos de linguagem (Large Language Models – LLMs). Criminosos utilizam Inteligência Artificial para manter conversas emocionalmente sofisticadas, persistentes e convincentes, eliminando erros linguísticos anteriormente associados a

perfis falsos. Essas interações simulam vínculos afetivos genuínos, explorando vulnerabilidades emocionais para obtenção de dinheiro, dados pessoais ou manipulação psicológica. Além disso, mulheres são frequentemente vítimas de ataques reputacionais envolvendo deepfakes, pornografia não consensual e manipulação audiovisual. Em uma sociedade profundamente marcada pela exposição digital da identidade, conteúdos falsificados possuem elevado potencial destrutivo sobre a vida pessoal, profissional e social das vítimas. A utilização indevida da imagem feminina em conteúdos sintéticos reforça estruturas históricas de violência simbólica e controle sobre corpos e identidades femininas. Muitas vítimas enfrentam consequências severas relacionadas à saúde mental, exclusão social, prejuízos profissionais e revitimização constante nas plataformas digitais. Nesse cenário, torna-se indispensável o fortalecimento de mecanismos jurídicos capazes de responsabilizar os autores dessas práticas e garantir maior proteção aos direitos da personalidade. A legislação precisa acompanhar a velocidade das transformações tecnológicas, reconhecendo a gravidade das violências digitais de gênero e ampliando instrumentos de proteção às vítimas.

## 5.4 O caminho para uma navegação mais Segura

Diante da expansão das fraudes digitais baseadas em Inteligência Artificial, torna-se necessário desenvolver estratégias estruturais capazes de promover ambientes digitais mais seguros, transparentes e eticamente responsáveis. Nesse contexto, ganha relevância o conceito de "Ética por Design", baseado na incorporação de princípios éticos e mecanismos preventivos desde a concepção das tecnologias digitais. Uma das medidas mais discutidas refere-se à implementação obrigatória de mecanismos de identificação de conteúdos produzidos por IA, como marcas d'água digitais, metadados verificáveis e sistemas de autenticação audiovisual. Essas ferramentas podem contribuir para diferenciar conteúdos autênticos de produções sintéticas, reduzindo impactos da desinformação e da manipulação digital. Paralelamente, torna-se imprescindível ampliar investimentos em educação digital e letramento midiático para diferentes grupos sociais. A formação crítica da população deve envolver desde crianças até idosos, promovendo competências relacionadas à segurança da informação, identificação de golpes e análise crítica de conteúdos digitais. Outro aspecto fundamental refere-se à responsabilização das grandes plataformas digitais. As chamadas Big Techs exercem papel central na circulação de informações e precisam assumir responsabilidades proporcionais ao

impacto social de suas tecnologias. Isso inclui investimentos em sistemas automatizados de detecção de conteúdos maliciosos, transparência algorítmica e mecanismos rápidos de denúncia e remoção de materiais fraudulentos. Portanto, a construção de uma navegação mais segura depende da articulação entre tecnologia, educação, regulação e responsabilidade institucional. Mais do que combater fraudes isoladas, trata-se de fortalecer uma cultura digital baseada em ética, transparência, proteção de direitos fundamentais e desenvolvimento crítico da sociedade contemporânea.

## 6. Os meios utilizados para aplicação dos golpes

Segundo a Teoria da Atividade Rotineira, o crime não ocorre de maneira aleatória, mas resulta da convergência entre três elementos fundamentais: um ofensor motivado, um alvo adequado e a ausência de mecanismos eficazes de proteção ou vigilância (COHEN; FELSON, 2010). No contexto contemporâneo, o avanço da Inteligência Artificial generativa amplia significativamente a capacidade operacional dos agentes criminosos, aumentando a sofisticação, escala e personalização dos golpes digitais. O crescimento exponencial das fraudes associadas a smartphones e ambientes digitais evidencia essa transformação. De acordo com o Indicador de Tentativas de Fraude da Serasa Experian, em 2022 ocorria no Brasil uma tentativa de fraude a cada oito segundos. Apenas no mês de junho daquele ano foram registradas mais de 322 mil tentativas de golpes em território nacional. Paralelamente, a Federação Brasileira de Bancos (FEBRABAN) aponta que, desde o período da pandemia de Covid-19, houve crescimento expressivo das fraudes baseadas em dispositivos móveis, envolvendo práticas como roubo de identidade, extorsão e violação de dados pessoais. Nesse cenário, os mecanismos utilizados pelos criminosos evoluem continuamente, combinando engenharia social, manipulação psicológica e ferramentas tecnológicas avançadas para aumentar a eficácia dos ataques.

## 6.1 Engenharia social e fraudes com phishing

A engenharia social e as técnicas de phishing figuram entre os mecanismos mais utilizados em fraudes digitais contemporâneas. Sua evolução acompanha o desenvolvimento das tecnologias de comunicação, especialmente o crescimento da utilização de e-mails, aplicativos de mensagens instantâneas e redes sociais como principais meios de interação digital. Os ataques de phishing utilizam comunicações fraudulentas contendo links maliciosos, anexos infectados ou

páginas falsas com o objetivo de induzir usuários ao compartilhamento de informações confidenciais (NUNES, 2019). Entretanto, essa prática não se limita ao uso de correio eletrônico, podendo ocorrer por meio de chamadas telefônicas, mensagens SMS, aplicativos de mensagens instantâneas, redes sociais e até ambientes de jogos online. Segundo Alharthi e Zargari (2021 apud BEZERRA; SILVA; MILLIAN, 2024), o principal objetivo dessas ações é enganar a vítima, conduzindo-a ao acesso de páginas fraudulentas que imitam serviços legítimos, como bancos, lojas virtuais ou plataformas governamentais, para obtenção de credenciais, senhas e dados pessoais.

### 6.1.1 Phishing por e-mail

O phishing por e-mail ocorre quando criminosos utilizam mensagens eletrônicas maliciosas para solicitar informações privadas, direcionar usuários para páginas falsas ou induzir o download de arquivos comprometidos (ALHARTHI; ZARGARI, 2021 apud BEZERRA; MILLIAN; SILVA, 2024). Esses e-mails frequentemente simulam comunicações institucionais legítimas, utilizando logotipos, layouts e linguagem semelhantes aos de empresas confiáveis. A utilização de gatilhos emocionais, como urgência, bloqueio de contas ou promoções falsas, reduz a capacidade crítica do usuário e aumenta as chances de sucesso do golpe.

### 6.1.2 Spear phishing

O spear phishing caracteriza-se como uma modalidade direcionada de ataque, voltada especificamente para indivíduos, empresas ou instituições previamente selecionadas. Diferentemente do phishing genérico, essa técnica utiliza informações detalhadas sobre a vítima – obtidas em redes sociais, bancos de dados vazados ou plataformas digitais – para construir mensagens altamente personalizadas (ALEROUD; ZHOU, 2017). O grau de personalização torna essas abordagens mais convincentes, aumentando significativamente a probabilidade de comprometimento das vítimas. Em ambientes corporativos, essa modalidade é frequentemente utilizada para obtenção de credenciais institucionais, espionagem digital e fraudes financeiras.

### 6.1.3 Smishing

O smishing corresponde ao phishing realizado por meio de mensagens SMS. Nessa modalidade, criminosos utilizam mensagens de texto para induzir usuários a acessar links fraudulentos, compartilhar informações sensíveis ou entrar em contato com centrais falsas de atendimento (ALHARTHI; ZARGARI, 2021 apud BEZERRA; MILLIAN; SILVA, 2024). No Brasil, ganhou destaque o chamado “golpe do 0800”, no qual vítimas recebem mensagens alegando movimentações suspeitas em contas bancárias ou compras indevidas. A mensagem orienta o usuário a entrar em contato com um número iniciado pelo prefixo 0800, explorando a percepção de credibilidade associada a serviços institucionais gratuitos (MIATO, 2024). Essa estratégia aumenta significativamente a confiança da vítima na comunicação recebida, ampliando a eficácia da fraude.

### 6.1.4 Vishing

O vishing (voice phishing) envolve a utilização de chamadas telefônicas fraudulentas para obtenção de informações confidenciais ou indução da vítima à realização de determinadas ações financeiras ou operacionais. Nessa modalidade, criminosos se passam por representantes de bancos, empresas, órgãos públicos ou familiares, utilizando técnicas de engenharia social para persuadir a vítima a compartilhar senhas, códigos de autenticação ou realizar transferências financeiras (BEZERRA; SILVA; MILLIAN, 2024). Com o avanço da Inteligência Artificial, o vishing passou a incorporar recursos de clonagem de voz e deepfakes sonoros, tornando as abordagens significativamente mais convincentes.

### 6.1.5 Phishing HTTP

O phishing HTTP baseia-se na criação de páginas falsas que replicam visualmente websites legítimos, como instituições bancárias, plataformas de comércio eletrônico e serviços digitais. Essas páginas fraudulentas reproduzem logotipos, layouts e elementos gráficos quase idênticos aos originais, dificultando a identificação da fraude pelos usuários (BEZERRA; SILVA; MILLIAN, 2024). Ao acessar essas páginas, as vítimas inserem espontaneamente credenciais de acesso, dados bancários e informações pessoais, que são imediatamente capturados pelos criminosos. A crescente sofisticação visual dessas plataformas torna sua identificação cada vez mais complexa, especialmente em dispositivos móveis.

## 6.2 IA generativa e deepfakes

A Inteligência Artificial generativa inaugura uma nova etapa das fraudes digitais contemporâneas ao possibilitar a criação de conteúdos ultrarrealistas em formatos de texto, imagem, áudio e vídeo. Segundo a Revista DCS (2025), essas tecnologias desafiam diretamente a percepção de autenticidade e verdade no ambiente digital. Originalmente utilizadas para manipulações simples de imagens e vídeos, as tecnologias de deepfake evoluíram rapidamente para sistemas capazes de produzir conteúdos sintéticos altamente sofisticados, simulando falas, expressões faciais e comportamentos inexistentes com elevado grau de realismo (PAGBANK, 2024). De acordo com Monastier (2024), a produção dessas montagens envolve o mapeamento detalhado de características faciais, permitindo sincronização precisa de movimentos labiais, olhos e expressões. Como consequência, a detecção dessas manipulações torna-se cada vez mais difícil tanto para usuários comuns quanto para ferramentas tradicionais de segurança digital. Os impactos dessas tecnologias tornaram-se particularmente visíveis em contextos políticos. Nas eleições brasileiras de 2018 e 2022, diversos candidatos foram alvo de vídeos manipulados atribuídos falsamente a fontes confiáveis, contribuindo para desinformação e polarização social antes mesmo de verificações oficiais (NEUBER, 2025).

### 6.2.1 Voice deepfake

O voice deepfake utiliza Inteligência Artificial para reproduzir artificialmente a voz de indivíduos reais a partir de pequenos fragmentos de áudio coletados em redes sociais, vídeos públicos ou mensagens de voz (PAGBANK, 2025). Após a clonagem, criminosos realizam chamadas ou enviam áudios simulando familiares, gestores ou representantes institucionais, solicitando transferências financeiras urgentes ou compartilhamento de códigos de autenticação. A credibilidade gerada pela reprodução da voz torna esse tipo de golpe extremamente eficaz, sobretudo em situações emocionalmente sensíveis.

### 6.2.2 Video deepfake

O vídeo deepfake corresponde à manipulação ou criação de vídeos sintéticos capazes de simular que determinada pessoa realizou ações ou pronunciou discursos inexistentes (PAGBANK, 2025). Essa técnica é utilizada para imitar executivos, autoridades públicas,

familiares ou celebridades em chamadas de vídeo e conteúdos compartilhados em redes sociais, aumentando significativamente a credibilidade da fraude e potencializando processos de desinformação.

### 6.2.3 Fraudes em tempo real em chamadas ao vivo

Com o avanço das ferramentas de IA generativa, criminosos passaram a utilizar filtros e sistemas de manipulação audiovisual em tempo real durante videoconferências. Essa modalidade permite assumir instantaneamente a aparência e voz de terceiros em reuniões corporativas ou contatos pessoais, possibilitando autorizações fraudulentas de pagamentos, acesso indevido a sistemas internos e obtenção de informações estratégicas. A manipulação em tempo real representa uma das formas mais sofisticadas de fraude digital contemporânea, desafiando mecanismos tradicionais de autenticação visual e sonora.

### 6.2.4 Deepfake em redes sociais

As redes sociais também se tornaram ambientes propícios para disseminação de deepfakes. Perfis falsos são criados utilizando imagens e vídeos manipulados para simular identidades reais, incluindo pessoas conhecidas, influenciadores digitais e figuras públicas. Esses perfis são utilizados para obtenção de confiança, disseminação de links maliciosos, aplicação de golpes financeiros e propagação de campanhas de desinformação. A combinação entre Inteligência Artificial, algoritmos de recomendação e viralização de conteúdo amplia significativamente o alcance e a velocidade dessas fraudes digitais.

## 6.3 Spoofing e fraudes digitais

O spoofing consiste na prática de mascarar comunicações digitais para fazê-las parecer provenientes de fontes legítimas e confiáveis (NASCIMENTO, 2020 apud GSI, 2019). Essa técnica pode envolver falsificação de números telefônicos, endereços IP, domínios, e-mails e identidades digitais. Segundo Morgado et al. (2023), o spoofing frequentemente integra estratégias de engenharia social utilizadas por organizações criminosas para aplicação massiva de golpes. Sistemas automatizados realizam milhares de contatos simultaneamente até identificar vítimas suscetíveis à manipulação emocional. Além das ligações telefônicas, o spoofing também

está associado a fraudes financeiras, clonagem de websites, manipulação de endereços eletrônicos e falsificação de identidades digitais. Sua principal finalidade consiste em criar aparência de legitimidade para aumentar a confiança das vítimas e facilitar obtenção de informações sensíveis. Entre os grupos mais vulneráveis destacam-se os idosos. O crescimento da participação dessa população no ambiente digital foi acompanhado pelo aumento expressivo de fraudes direcionadas especificamente a esse público (MORGADO et al., 2023). Segundo Morgado (2024), fatores como menor letramento digital, confiança em contatos institucionais e desconhecimento sobre mecanismos de fraude tornam pessoas idosas especialmente suscetíveis às estratégias de engenharia social. Os criminosos exploram emoções, sobrecarga de informações e relações de confiança para influenciar processos decisórios, utilizando falsas promessas, senso de urgência e construção artificial de vínculos afetivos como mecanismos de manipulação psicológica (MORGADO et al., 2023). Dessa forma, observa-se que os meios utilizados para aplicação dos golpes digitais combinam vulnerabilidades técnicas, manipulação emocional e tecnologias emergentes baseadas em Inteligência Artificial, configurando um dos principais desafios contemporâneos da segurança cibernética e da proteção de dados pessoais.

## 7. Estratégias de prevenção e mitigação de riscos digitais

A consolidação da sociedade hiperconectada transformou dispositivos móveis, redes digitais e plataformas online em extensões permanentes da própria experiência humana. Smartphones, aplicativos e sistemas em nuvem passaram a centralizar informações pessoais, registros financeiros, memórias afetivas, relações sociais e atividades profissionais, tornando os indivíduos continuamente expostos aos ambientes digitais. Entretanto, essa integração tecnológica não foi acompanhada por uma evolução proporcional da cultura de segurança digital da população. Embora ameaças tradicionais, como phishing, engenharia social e roubo de credenciais, já representassem desafios significativos, o cenário contemporâneo tornou-se ainda mais complexo diante da convergência entre Inteligência Artificial generativa e manipulação psicológica. A combinação entre IA e engenharia social potencializa a criação de ataques altamente personalizados, sofisticados e emocionalmente convincentes. Ferramentas capazes de produzir deepfakes, clonagem de voz, mensagens automatizadas e conteúdos sintéticos ultrarrealistas ampliam significativamente a vulnerabilidade dos usuários, dificultando a identificação de fraudes mesmo por indivíduos experientes. Nesse contexto, a vulnerabilidade

digital deixa de ser exclusivamente técnica e passa a assumir dimensões cognitivas, emocionais e sociais. O usuário torna-se alvo prioritário das ofensivas cibernéticas não apenas por limitações operacionais, mas pela exploração sistemática de emoções humanas como medo, urgência, confiança e afeto. A superexposição digital, associada à velocidade de circulação das informações e à dependência crescente de ambientes virtuais, contribui para um cenário no qual a manipulação informacional se torna cada vez mais invisível e naturalizada. Assim, compreender a fragilidade mental e social dos indivíduos diante das tecnologias emergentes torna-se elemento central para o desenvolvimento de estratégias eficazes de prevenção e mitigação de riscos digitais.

## 7.1 A cibersegurança como ferramenta de proteção

Embora frequentemente utilizados como sinônimos, os conceitos de segurança da informação e cibersegurança possuem distinções relevantes. A segurança da informação concentra-se prioritariamente na proteção dos dados enquanto ativos estratégicos, abrangendo princípios como confidencialidade, integridade e disponibilidade das informações. A cibersegurança, por sua vez, amplia essa perspectiva ao considerar não apenas a proteção técnica dos sistemas digitais, mas também a defesa dos indivíduos inseridos nesses ambientes. Nesse contexto, o usuário deixa de ser visto apenas como operador de sistemas e passa a ser compreendido como alvo central das ameaças digitais contemporâneas (REEGÅRD; BLACKETT; KATTA apud MATHEWS, 2025). Segundo Taddeo (apud MATHEWS, 2025), a cibersegurança deve ser compreendida como um bem público essencial para a estabilidade da sociedade da informação. Essa abordagem parte do reconhecimento de que as infraestruturas digitais são intrinsecamente vulneráveis, exigindo estratégias coletivas e permanentes de proteção. A crescente digitalização dos serviços financeiros, sistemas governamentais, hospitais, plataformas educacionais e infraestruturas críticas evidencia que ataques cibernéticos não afetam apenas indivíduos isolados, mas podem comprometer setores inteiros da sociedade. Em razão disso, instituições públicas e privadas vêm ampliando investimentos em mecanismos de proteção digital, monitoramento de ameaças e políticas de segurança da informação. Bancos, Big Techs, operadoras de telecomunicações e instituições de saúde figuram entre os setores mais sensíveis aos impactos das fraudes digitais, sobretudo devido ao elevado volume de dados pessoais e financeiros armazenados em seus sistemas. Nesse cenário, a cibersegurança passa a ocupar

posição estratégica não apenas no campo tecnológico, mas também nas dimensões econômica, política e social contemporâneas.

## 7.2 Vetores de ataque

Considerando que o indivíduo se tornou o principal alvo das fraudes digitais contemporâneas, as estratégias de defesa precisam ultrapassar barreiras puramente tecnológicas e incorporar dimensões comportamentais, educativas e cognitivas. Mesmo sistemas avançados de proteção tornam-se insuficientes quando usuários são manipulados emocionalmente ou induzidos a fornecer voluntariamente informações sensíveis. A exploração do fator humano consolidou-se como um dos principais vetores de ataque utilizados por criminosos digitais. Nesse contexto, mecanismos tradicionais de defesa baseados apenas em antivírus, firewalls ou autenticação técnica precisam ser complementados por estratégias contínuas de conscientização, educação digital e fortalecimento do senso crítico dos usuários. Além disso, a evolução da Inteligência Artificial ampliou significativamente a capacidade de personalização dos ataques. Criminosos conseguem combinar informações obtidas em redes sociais, bancos de dados vazados e ferramentas automatizadas para produzir abordagens altamente convincentes, dificultando mecanismos intuitivos de identificação de fraudes. Dessa forma, as estratégias contemporâneas de mitigação de riscos exigem integração entre tecnologias de detecção, políticas regulatórias, educação digital e mecanismos permanentes de verificação da autenticidade das comunicações digitais.

## 7.3 Spoofing

O spoofing constitui um dos principais vetores utilizados em fraudes digitais contemporâneas. Essa técnica baseia-se na falsificação de identidades digitais – como números telefônicos, endereços de e-mail, domínios ou endereços IP – para fazer com que comunicações fraudulentas aparentam origem legítima. A eficácia do spoofing decorre principalmente da exploração da confiança dos usuários em instituições reconhecidas, como bancos, operadoras telefônicas, órgãos governamentais e empresas privadas. Ao acreditar que está interagindo com uma fonte confiável, a vítima tende a reduzir mecanismos naturais de desconfiança e compartilhar informações sensíveis com maior facilidade. Com o avanço das tecnologias de IA generativa, o spoofing passou a incorporar recursos sofisticados de clonagem de voz e

manipulação audiovisual em tempo real, tornando os golpes significativamente mais convincentes. Chamadas telefônicas fraudulentas utilizando vozes sintéticas de familiares ou representantes institucionais passaram a integrar o repertório contemporâneo da engenharia social. Diante desse cenário, a mitigação desse tipo de ameaça exige múltiplas estratégias complementares. Entre as principais medidas preventivas destacam se: • utilização de autenticação em dois fatores (2FA); • verificação da identidade do interlocutor por canais alternativos; • desconfiança de solicitações urgentes envolvendo transferências financeiras; • atualização constante de sistemas e aplicativos; • utilização de mecanismos de autenticação de chamadas; • fortalecimento da educação digital e da conscientização dos usuários. Além disso, especialistas defendem maior rigor regulatório sobre operadoras de telecomunicações e plataformas digitais, bem como ampliação de investimentos em sistemas automatizados de detecção de padrões fraudulentos baseados em Inteligência Artificial. Assim, o enfrentamento do spoofing e demais vetores de ataque contemporâneos depende da articulação entre tecnologia, educação, regulamentação e desenvolvimento de uma cultura digital baseada em prevenção, verificação e consciência crítica.

## Conclusão

Diante do cenário exposto ao longo deste estudo, observa-se que a evolução da Inteligência Artificial generativa não apenas sofisticou as fraudes digitais contemporâneas, mas também evidenciou de forma contundente as vulnerabilidades estruturais, sociais e psicológicas presentes na sociedade hiperconectada. O aprimoramento de práticas como phishing, spoofing, engenharia social e deepfakes demonstra que os crimes cibernéticos deixaram de depender exclusivamente de conhecimentos técnicos avançados, passando a explorar, de maneira estratégica, emoções humanas, relações de confiança e fragilidades cognitivas. Nesse contexto, a Inteligência Artificial transforma-se em instrumento capaz de potencializar mecanismos de manipulação psicológica altamente personalizados, ampliando significativamente o alcance e a eficácia das fraudes digitais. Os impactos dessas práticas ultrapassam as perdas financeiras imediatas, atingindo dimensões relacionadas à confiança social, à estabilidade informacional, à reputação individual e à saúde mental das vítimas. A disseminação de conteúdos sintéticos, associada à crescente dificuldade de distinguir o real do artificial, contribui para a formação de um ambiente de insegurança permanente, no qual usuários passam a desconfiar continuamente

das informações, imagens, vídeos e comunicações que consomem no cotidiano digital. Esse cenário afeta diretamente grupos socialmente mais vulneráveis, como idosos, mulheres, crianças e adolescentes, que frequentemente se tornam alvos prioritários de golpes baseados em manipulação emocional, violência de gênero digital e exploração de vínculos afetivos. Além disso, observa-se que a velocidade de evolução das tecnologias criminosas supera, em muitos casos, a capacidade de adaptação das estruturas jurídicas e institucionais. Embora existam avanços normativos importantes no campo da proteção de dados e do combate aos crimes cibernéticos, a legislação ainda enfrenta dificuldades para acompanhar a sofisticação das ameaças baseadas em IA generativa, especialmente no que se refere à responsabilização, rastreamento e validação de provas digitais. Dessa forma, a mitigação dos riscos impostos pela Inteligência Artificial exige uma abordagem multidimensional, capaz de integrar tecnologia, regulação, educação e fortalecimento da cultura de segurança digital. O enfrentamento dessas ameaças não pode restringir-se apenas ao desenvolvimento de ferramentas técnicas de proteção, sendo igualmente necessário investir na formação crítica da sociedade e na ampliação do letramento digital. Nesse sentido, a educação digital emerge como uma das principais linhas de defesa contra fraudes contemporâneas, permitindo o desenvolvimento de competências relacionadas à verificação de informações, identificação de manipulações e construção de uma postura crítica diante das interações digitais. Paralelamente, torna-se indispensável a criação de mecanismos regulatórios mais robustos, políticas públicas de proteção digital e instrumentos tecnológicos capazes de identificar e mitigar ataques baseados em Inteligência Artificial. Por fim, compreende-se que a preservação da dignidade humana no ambiente digital depende da construção de um equilíbrio entre inovação tecnológica, responsabilidade ética e proteção dos direitos fundamentais. Somente por meio da articulação entre educação, regulamentação, cibersegurança e conscientização coletiva será possível garantir que a tecnologia permaneça a serviço da sociedade, e não como instrumento de ampliação das vulnerabilidades e desigualdades contemporâneas.

# REFERÊNCIAS

**Referências Organizadas (ABNT)**

TIC CENTRO REGIONAL DE ESTUDOS PARA O DESENVOLVIMENTO DA SOCIEDADE DA INFORMAÇÃO (CETIC). *Domicílios 2023*. Disponível em: CETIC.br. Acesso em: 14 abr. 2026.

AGÊNCIA PÚBLICA. *Deep nudes: vítimas enfrentam impunidade e sites lucram com IA*. 2025.

ALEROUD, Ahmed; ZHOU, Lijun. Phishing environments, techniques, and countermeasures. *Computers & Security*, v. 68, p. 160–196, 2017. DOI: 10.1016/j.cose.2017.05.006.

ALHARTHI, Sultan; ZARGARI, Shayan. Phishing attacks and countermeasures. In: *UIC 2021*. [S.l.]: IEEE, 2021. DOI: 10.1109/UIC52205.2021.00019.

ANASTACIO, T. F.; TAVARES, C. Q. Imagem roubada: o uso de deepfakes durante o período de campanha eleitoral de 2024. In: SIMPÓSIO DE COMUNICAÇÃO DA REGIÃO TOCANTINA (SIMCUM), 18., 2024, Maranhão. *Anais [...]* Maranhão: UFMA, 2024.

AZEVEDO, Ingrid Borges de. *Deepfakes contra candidatas nas eleições*. Brasília: Universidade de Brasília, 2025.

BATISTA, A.; SANTAELLA, L. Desinformação, deepfakes e os impactos no sistema eleitoral. *Organicom*, São Paulo, 2024.

BAUMAN, Zygmunt Bauman. *Modernidade líquida*. Rio de Janeiro: Zahar, 2001.

BEZERRA NETO, Heriberto Escolástico; SIQUEIRA, Mariana de. Dados na internet: uma análise das fraudes virtuais. *Revista FOCO*, v. 18, n. 11, 2025.

BEZERRA, P. S.; MILLIAN, L. E. A.; SILVA, R. C. Estudo de técnicas de phishing: métodos de ataque e estratégias de defesa. *Revista Mackenzie*, 2024.

BEZERRA, Paloma de Sousa; SILVA, R. C.; MILLIAN, L. E. A. Estudo de técnicas de phishing: métodos de ataque e estratégias de defesa. 2024. DOI: 10.55905/revconv.18n.7-074.

BRASIL. Lei nº 14.478, de 21 de dezembro de 2022. Dispõe sobre prestadores de serviços de ativos virtuais. *Diário Oficial da União*, Brasília, DF, 22 dez. 2022.

COHEN, Lawrence E.; FELSON, Marcus. Social change and crime rate trends: a routine activity approach. *American Sociological Review*, v. 44, n. 4, p. 588–608, 1979. DOI: 10.2307/2094589.

FEBRABAN – FEDERAÇÃO BRASILEIRA DE BANCOS. *Golpes digitais e IA: relatório de segurança 2026*. Disponível em: FEBRABAN. Acesso em: 9 maio 2026.

LÉVY, Pierre Lévy. *Cibercultura*. São Paulo: Editora 34, 1999.

LUCAS, Edmar Luiz. *Manipulação digital em rede: deepfakes, bots e milícias virtuais como desafios contemporâneos à segurança da informação*. 2025. Projeto Monográfico (Segurança da Informação) – Fatec Americana, Americana, 2025.

MACEDO, R. C.; SILVA, M. F. P. O golpe está aí: vulnerabilidade social e prevenção dos crimes phishing. *JNT-Facit Business and Technology Journal*, 2025.

MARTINS, Sanielly Aurea de Souza. *Deepfake e a violação ao direito da personalidade: análise da percepção de pessoas na cidade de Juazeiro do Norte-CE*. 2024. Trabalho de Conclusão de Curso (Direito) – UNILEÃO, Juazeiro do Norte, 2024.

MATHEWS, J. R. S. Análise dos principais golpes no Brasil e métodos de prevenção. *Boletim de Fraudes Digitais*, 2025.

MATHEWS, Juan Mathews Rebello Santos. *Análise dos principais golpes no Brasil e métodos de prevenção*. Relatório Técnico, set. 2025.

MIATO. Golpe do 0800: criminosos se passam por bancos para roubar dados. *InfoMoney*, 2024. Disponível em: InfoMoney. Acesso em: 9 maio 2026.

MOLINA, Adriano Cezar; BERENGUEL, Orlando Leonardo. Deepfake: a evolução das fake news. *Research, Society and Development*, v. 11, n. 6, 2022.

MONASTIER, Nilton Cesar Kleina. Deepfake e golpes: como se proteger de cibercriminosos? *TecMundo*, 2024. Disponível em: TecMundo. Acesso em: 9 maio 2026.

MORGADO, Eduardo Martins et al. Caso de cyber fraude por telefone no Brasil e a inteligência artificial: vítimas idosas, spoofing até a manipulação por engenharia social. In: *Inteligência artificial e suas aplicações interdisciplinares*. [S.l.]: Editora e-Publicar, 2023. p. 113-126. DOI: 10.47402/ed.ep.c202321007201.

NASCIMENTO, Adeilson Evangelista. *Spoofing nas redes de telefonia: alternativas para mitigação*. 2020. Trabalho de Conclusão de Curso (Especialização) – Escola Superior de Guerra, Brasília, 2020.

NEGREIRO, Rodrigo. *Deepfakes e confiança digital na era da inteligência artificial*. [S.l.]: [s.n.], 2025.

NEMER, David Nemer. *Tecnologia do oprimido: desigualdade e apropriação tecnológica nas periferias*. São Paulo: [s.n.], 2021.

NEUBER, Diego. Desinformação e deepfakes como vetores emergentes de ameaças cibernéticas no Brasil. *Revista Latino-Americana de Estudos Científicos*, v. 6, n. 27, 2025. DOI: 10.55470/relaec.48740.

O'NEIL, Cathy O'Neil. *Armas de destruição matemática: como o Big Data aumenta a desigualdade e ameaça a democracia*. New York: Crown, 2020.

PAGBANK. Deepfake: como funciona a nova modalidade de golpe com inteligência artificial. 2024. Disponível em: PagBank Blog. Acesso em: 9 maio 2026.

POLIDO, Gabriel Carvalho. *Estudo sobre fraudes digitais e o desenvolvimento de aplicativo para smartphones*. 2023. Trabalho de Conclusão de Curso (Bacharelado) – UNESP, Bauru, 2023.

REVISTA DCS. Deepfakes e responsabilidade penal: novos desafios para o direito penal digital. *Revista DCS*, v. 22, n. 83, 2025. DOI: 10.54899/dcs.v22i84.3604.

SAFERNET BRASIL. *Indicadores de crimes cibernéticos no Brasil*. Disponível em: SaferNet Brasil. Acesso em: 12 maio 2026.

SERASA EXPERIAN. Pessoas entre 36 e 50 anos são os principais alvos de golpistas, aponta pesquisa da Serasa Experian. 2022. Disponível em: Serasa Experian. Acesso em: 9 maio 2026.

SOARES, B. B.; RIBEIRO FILHO, C. F. *Cibersegurança: ameaças de phishing relacionadas a roubo de identidade*. Varginha: UNIS, 2022.

SOUSA, Gesson Eliésio Aguiar de et al. O impacto da inteligência artificial generativa na sofisticação de fraudes e golpes online com uso de deepfakes no Brasil. *Revista DCS*, v. 22, n. 84, 2025.

SOUZA, A. D.; RODRIGUES, P. S.; FERREIRA, R. M. Fraude digital e phishing: desafios para a eficácia da lei penal. *Revista Multidebates*, 2025.

SPINOLA, Luíza Moura Costa. *O tratamento do spoofing conforme a legislação penal brasileira*. 2020. Dissertação (Mestrado em Direito) – UCSAL, Salvador, 2020.

SUMSUB. *Identity Fraud Report 2024*. [S.l.]: Sumsub, 2024.

UNIÃO INTERNACIONAL DE TELECOMUNICAÇÕES (ITU). *Facts and Figures 2021*. Disponível em: ITU Facts and Figures 2021. Acesso em: 14 abr. 2026.