V. 06, N.27 Jan./Jun. 2025

DESINFORMAÇÃO E DEEPFAKES COMO VETORES EMERGENTES DE AMEAÇAS CIBERNÉTICAS NO BRASIL

DISINFORMATION AND DEEPFAKES AS EMERGING VECTORS OF CYBER THREATS IN BRAZIL

LA DESINFORMACIÓN Y LOS DEEPFAKES COMO VECTORES EMERGENTES DE AMENAZAS CIBERNÉTICAS EN BRASIL

Diego Neuber
Universidade Norte do Paraná
ORCID – https://orcid.org/0009-0001-6474-5218

Resumo: A crescente sofisticação de tecnologias de inteligência artificial, especialmente as voltadas à criação de conteúdo sintético, tem impulsionado uma nova geração de ameaças cibernéticas. Entre elas, destacam-se a desinformação digital e os deepfakes — conteúdos manipulados que simulam, com alto grau de realismo, rostos, vozes e ações humanas. Este artigo investiga como esses recursos têm sido utilizados como vetores de ataques no contexto brasileiro, afetando processos eleitorais, a confiança em instituições públicas, a reputação de empresas e a segurança das informações. A partir de revisão de literatura técnica, relatórios de segurança, estudos de caso nacionais e internacionais e dados empíricos, propõe-se uma análise crítica do impacto dessas tecnologias na esfera cibernética. Ao final, são discutidas medidas estratégicas de mitigação que envolvem alfabetização digital, detecção automatizada, regulamentação e governança ética da IA.
Palavras-chave: Cibersegurança. Desinformação. Deepfakes. Inteligência Artificial. Engenharia Social. Risco digital.

Abstract: The growing sophistication of artificial intelligence technologies, especially those aimed at creating synthetic content, has driven a new generation of cyber threats. These include digital disinformation and deepfakes - manipulated content that simulates human faces, voices and actions with a high degree of realism. This article investigates how these resources have been used as attack vectors in the Brazilian context, affecting electoral processes, trust in public institutions, the reputation of companies and information security. Based on a review of technical literature, security reports, national and international case studies and empirical data, it proposes a critical analysis of the impact of these technologies in the cyber sphere. Finally, strategic mitigation measures involving digital literacy, automated detection, regulation and ethical governance of AI are discussed.
Keywords: Cybersecurity. Disinformation. Deepfakes. Artificial Intelligence. Social engineering. Digital risk.

Resumen: La creciente sofisticación de las tecnologías de inteligencia artificial, especialmente las destinadas a crear contenidos sintéticos, ha alimentado una nueva generación de ciberamenazas. Entre ellas se encuentran la desinformación digital y los deepfakes -contenidos manipulados que simulan rostros, voces y acciones humanas con un alto grado de realismo-. Este artículo investiga cómo estos recursos han sido utilizados como vectores de ataque en el contexto brasileño, afectando a los procesos electorales, la confianza en las instituciones públicas, la reputación de las empresas y la seguridad de la información. A partir de una revisión de literatura técnica, informes de seguridad, estudios de casos nacionales e internacionales y datos empíricos, se propone un análisis crítico del impacto de estas tecnologías en la ciberesfera. Por último, se discuten medidas estratégicas de mitigación que incluyen la alfabetización digital, la detección automatizada, la regulación y la gobernanza ética de la IA.
Palabras-clave: Ciberseguridad. Desinformación. Deepfakes. Inteligencia Artificial. Ingeniería social. Riesgo digital.

INTRODUÇÃO

Nos últimos anos, o Brasil testemunhou uma transformação no cenário de ameaças cibernéticas. De ataques puramente técnicos, como infecções por malware e ransomware, evoluiu-se para estratégias mais complexas, nas quais a manipulação da informação tornou-se arma central. O fenômeno da desinformação digital, aliado ao avanço das tecnologias de inteligência artificial generativa, resultou em uma combinação poderosa e perigosa: os deepfakes.

Deepfakes são conteúdos sintéticos — vídeos, áudios ou imagens — criados por redes neurais capazes de simular com alta fidelidade expressões faciais, entonações de voz e movimentos corporais. Embora possuam usos legítimos, como no cinema e na educação, também vêm sendo utilizados para fins maliciosos, especialmente em contextos políticos e empresariais. No Brasil, casos envolvendo vídeos falsificados de figuras públicas circularam amplamente nas eleições de 2018 e 2022, gerando instabilidade, polarização e desinformação em larga escala.

Este artigo busca analisar criticamente o uso de desinformação e deepfakes como vetores de ataque cibernético, compreender os impactos

sociais e institucionais dessa prática e propor abordagens concretas de mitigação voltadas ao contexto nacional.

FUNDAMENTAÇÃO TEÓRICA

A desinformação pode ser definida como o compartilhamento intencional de informações falsas com o objetivo de enganar, influenciar ou manipular decisões sociais, políticas ou econômicas (Wardle e Derakhshan, 2017). Ao contrário da mera fake news, a desinformação estruturada opera com narrativas coordenadas, uso estratégico de algoritmos e, cada vez mais, com apoio de IA para amplificação de conteúdo.

Os deepfakes, por sua vez, são uma evolução da manipulação digital tradicional. Criados a partir de Redes Adversárias Generativas (GANs), esses conteúdos podem recriar a imagem ou voz de uma pessoa com tamanha precisão que se tornam quase indistinguíveis do real (MIT Technology Review, 2023). Do ponto de vista da cibersegurança, os deepfakes representam um novo tipo de ameaça baseada em engenharia social avançada, com capacidade de enganar até sistemas biométricos (Sensity AI, 2022).

De acordo com relatório da ENISA (2021), a desinformação habilitada por IA é considerada uma das principais ameaças emergentes na União Europeia, sobretudo quando empregada para manipulação de eleições, fraudes corporativas e ataques à credibilidade de instituições públicas.

O CENÁRIO BRASILEIRO: FRAGILIDADE INFORMACIONAL E ALTA EXPOSIÇÃO

O Brasil é particularmente vulnerável à desinformação digital por uma série de fatores estruturais e culturais. Segundo o Datafolha (2022), mais de 79% dos brasileiros acessam as redes sociais como principal fonte de informação — muitas vezes sem realizar checagens de veracidade. A baixa educação midiática somada à polarização política extrema cria um

ambiente fértil para a propagação de boatos, distorções e, mais recentemente, vídeos manipulados com aparência legítima.

As eleições de 2018 e 2022 exemplificam esse fenômeno. Diversos candidatos foram alvos de vídeos falsificados, atribuídos a fontes supostamente confiáveis, com conteúdos difamatórios ou distorcidos. Em muitos casos, esses materiais circularam antes que as autoridades eleitorais conseguissem desmenti-los. A lentidão da verificação institucional, associada à velocidade dos compartilhamentos nas redes, favorece a eficácia desse tipo de ataque.

Além da política, o ambiente corporativo também tem sido afetado. Empresas de médio e grande porte relataram casos de fraudes por voz simulada, nas quais deepfakes de executivos solicitavam transferências ou vazamento de dados sensíveis (IBM X-Force, 2023).

DEEPFAKES COMO VETOR DE AMEAÇA CIBERNÉTICA

A atuação dos deepfakes como vetor de ataque ocorre de maneira transversal. A seguir, destacam-se os principais impactos:

- **Fraudes Financeiras**

  Golpistas têm utilizado deepfakes de voz e vídeo para se passar por diretores de empresas, solicitando transações fraudulentas a funcionários do setor financeiro. Em 2021, uma empresa britânica perdeu € 220 mil em um golpe desse tipo (Forbes, 2022).

- **Ataques à Reputação**

  Conteúdos forjados com imagens de figuras públicas ou executivos de empresas têm sido utilizados para descredibilizá-los ou vincular suas imagens a discursos falsos. Tais ataques são difíceis de reverter, mesmo após a comprovação da falsidade do material.

- **Comprometimento de Dados e Privacidade**

  Deepfakes têm sido usados para simular reuniões, obter credenciais de acesso e burlar autenticação por reconhecimento facial e voz. Isso representa um risco crescente para a segurança de ambientes corporativos e infraestruturas críticas.

PROPOSTAS DE MITIGAÇÃO E RESPOSTA ESTRATÉGICA

A mitigação dos riscos associados à desinformação e aos deepfakes requer ação multidimensional — combinando tecnologia, educação, política pública e colaboração intersetorial.

- **Educação e Alfabetização Digital**

  É imperativo investir na formação crítica da população quanto ao consumo de informações. Isso inclui:

  Inclusão de **educação midiática** nos currículos escolares

  Capacitação de profissionais da educação para ensinar verificação de fatos

  Campanhas públicas regulares sobre reconhecimento de conteúdo falso

- **Monitoramento Automatizado e Tecnologias de Detecção**

  Ferramentas como **Sensity AI**, **Deepware Scanner** e soluções baseadas em IA podem identificar padrões típicos de manipulação audiovisual. É crucial:

  -Integrar essas tecnologias a sistemas de segurança institucional

  -Adotar processos de verificação contínua em redes sociais e canais oficiais

  -Criar hubs nacionais de checagem automatizada em tempo real

- **Políticas de Segurança da Informação nas Organizações**

As empresas devem atualizar suas políticas de segurança, incluindo:
-Planos de resposta a incidentes envolvendo conteúdo falso
-Treinamento de porta-vozes para cenários de crise reputacional
-Simulações de ataques com engenharia social baseada em IA

- **Regulação e Governança Ética da Inteligência Artificial**
  É necessário avançar em marcos legais que:
  -Criminalizem o uso malicioso de deepfakes
  -Exijam **marcadores digitais de autenticidade** (ex: metadados certificados)
  -Incentivem **transparência algorítmica** em plataformas digitais

CONCLUSÃO

A desinformação e os deepfakes representam um novo paradigma de risco digital, em que os ataques não ocorrem mais apenas contra sistemas, mas contra a confiança humana. Em um país com vulnerabilidades estruturais como o Brasil, os danos potenciais vão além do individual, afetando a democracia, a economia e a estabilidade institucional. Enfrentar esse cenário exige uma resposta integrada, que una tecnologia de ponta, educação crítica e governança eficaz. A luta contra a desinformação, portanto, é também uma frente estratégica de cibersegurança.

REFERÊNCIAS

BRASIL. Tribunal Superior Eleitoral (TSE). Relatórios das Eleições 2018 e 2022.

CISA. Deepfakes and Synthetic Media Threats. U.S. Cybersecurity & Infrastructure Security Agency, 2022.

DATAFOLHA. "Desinformação e Redes Sociais no Brasil". 2022.

ENISA. Threat Landscape Report 2021–2022. European Union Agency for Cybersecurity.

FORBES. "Voice Deepfake Used to Scam Company Out of $243,000". 2022.

IBM X-Force. Threat Intelligence Index. IBM, 2023.

MIT TECHNOLOGY REVIEW. "The Deepfake Detection Arms Race". 2023.

SENSITY.AI. "Global Deepfake Threat Landscape 2022".

WARDLE, C.; DERAKHSHAN, H. Information Disorder. UNESCO Report, 2017