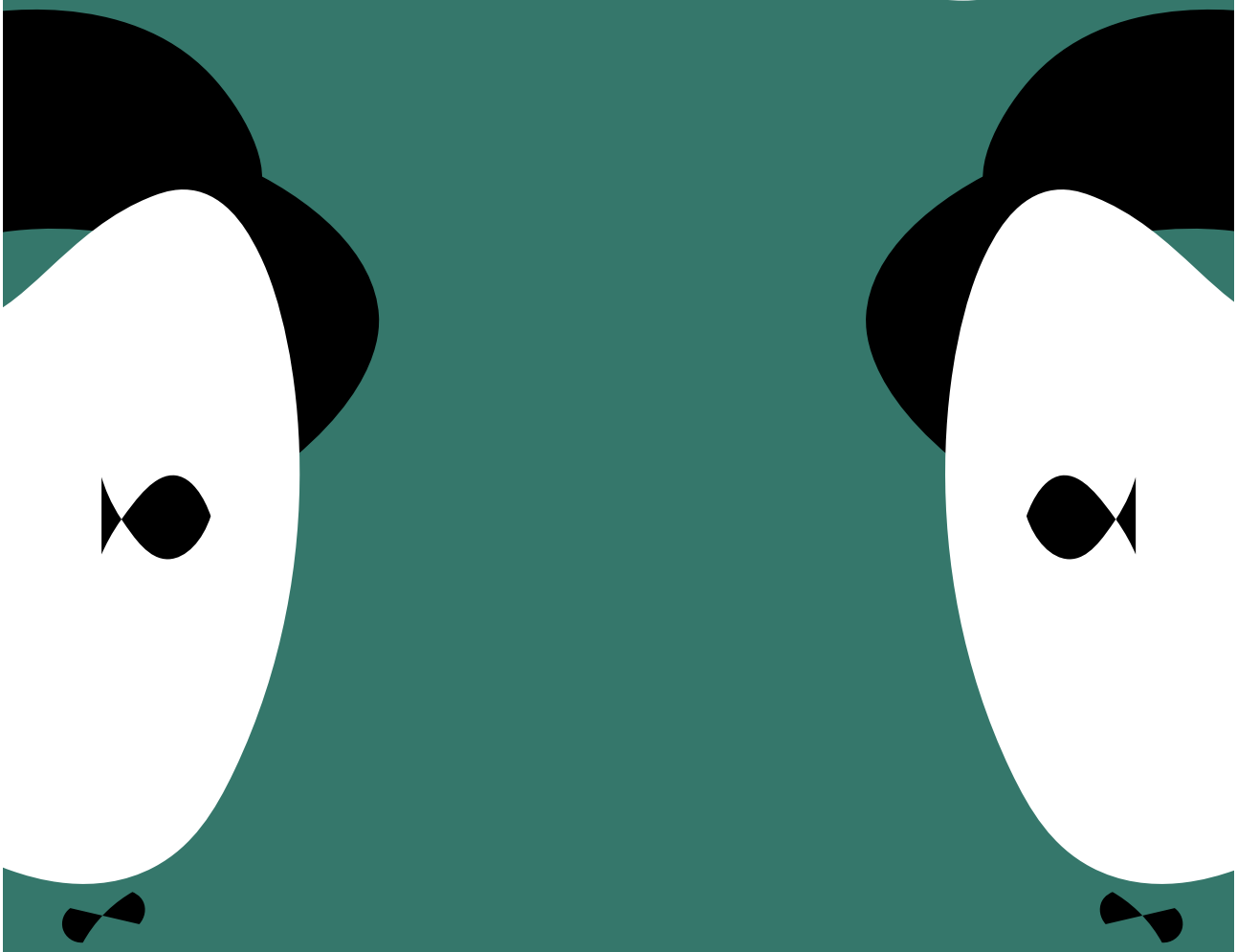
GUIA ILUSTRADO
CONTRA AS DEEPFAKES



STF
SUPREMO TRIBUNAL FEDERAL

DataPrivacyBR
Research

PROGRAMA DE COMBATE À
DESINFORMAÇÃO
SOCIEDADE INFORMADA, DEMOCRACIA FORTE

SUPREMO TRIBUNAL FEDERAL

Ministro **Luís Roberto Barroso**
Presidente 26.6.2013

Ministro **Luiz Edson Fachin**
Vice-Presidente 16.6.2015

Ministro **Gilmar Ferreira Mendes**
Decano 20.6.2002

Ministra **Cármen Lúcia Antunes Rocha**
21.6.2006

Ministro **José Antonio Dias Toffoli**
23.10.2009

Ministro **Luiz Fux**
3.3.2011

Ministro **Alexandre de Moraes**
22.3.2017

Ministro **Kassio Nunes Marques**
5.11.2020

Ministro **André Luiz de Almeida Mendonça**
16.12.2021

Ministro **Cristiano Zanin Martins**
3.8.2023

Ministro **Flávio Dino**
22.2.2024

FICHA TÉCNICA

Supremo Tribunal Federal

Secretária-Geral da Presidência
Aline Rezende Peres Osorio

Secretária de Relações com a Sociedade
Teresa Cristina de Melo Costa

Coordenadoria de Relações com a Sociedade e Combate à Desinformação
Victor Carnevalli Durigan
Frederico Franco Alvim
Marco Antonio Konopacki

Data Privacy Brasil

Diretoria
Bruno Bioni
Mariana Rielli
Rafael Zanatta

Autoria
Frederico Franco Alvim
Victor Carnevalli Durigan

Revisão técnica
Mariana Rielli
Marco Antonio Konopacki

Design e diagramação
Renato Barros de Carvalho
Rafael Regatieri

Capa e ilustrações
Renato Barros de Carvalho

Produção
Programa de Combate à Desinformação do Supremo Tribunal Federal e Data Privacy Brasil

COMO REFERENCIAR:

Brasil. Guia Ilustrado Contra as Deepfakes. Supremo Tribunal Federal; Data Privacy Brasil. Brasília: STF, Coordenadoria de Combate à Desinformação, 2024.

SUMÁRIO

INTRODUÇÃO

As tecnologias contemporâneas, como a inteligência artificial, possibilitam inúmeros avanços em favor da sociedade, em campos importantes como a medicina, a indústria, a educação e o sistema de justiça. Porém, essas aplicações são por vezes utilizadas por pessoas mal-intencionadas para aplicar golpes ou manchar a reputação de pessoas, grupos ou instituições.

DESINFORMAÇÃO

A evolução da tecnologia permite, como consequência, o desenvolvimento de novas modalidades de desinformação. Ferramentas de processamento de linguagem natural, por exemplo, podem dar vida a robôs (bots) programados para atuar, nas redes sociais, como usuários de carne-e-osso, com o fim de poluir, dificultar ou influenciar as discussões. Do mesmo modo, a inteligência artificial generativa pode ser usada para criar histórias fantasiosas, de forma coerente e com argumentos bastante convincentes, tanto em formato de texto como imagens e vídeos sintéticos muito realistas.

DEEPFAKES

Entre tantas novidades, uma delas se destaca: a tecnologia de produção das chamadas deepfakes. Especialistas consideram que esse tipo de conteúdo, em função da qualidade e do grau de sofisticação, representa riscos à saúde do debate público, a direitos individuais, como a honra, a imagem e o acesso a informações adequadas, e à normalidade de processos sociais sensíveis, como as eleições.

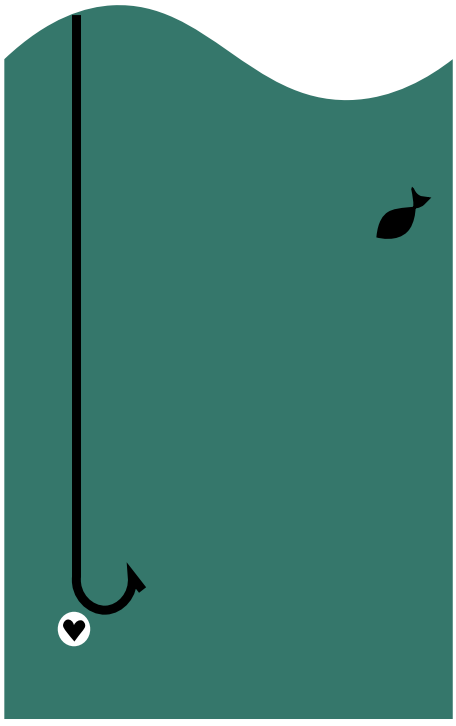
GUIA

Diante desta perspectiva, este Guia apresenta, em uma linguagem facilitada, informações essenciais para a plena compreensão do problema, incluindo instruções para que as pessoas possam identificar, evitar e denunciar a circulação de deepfakes.

O QUE SÃO DEEPFAKES?

A expressão "*deepfake*" surge da união dos termos "deep" – extraída da tecnologia *deep learning*, "aprendizado profundo" – e "*fake*", que significa "falso", em inglês. Não existe uma palavra em português para descrever esse fenômeno. Contudo, em tradução livre as deepfakes nada mais são do que "falsidades profundas", ou seja, conteúdos falsos produzidos com um alto grau de elaboração.

MULTIMÍDIA



No caso das deepfakes, a inteligência artificial é usada para gerar imagens, áudios ou vídeos fraudulentos, a partir da adulteração de elementos visuais (troca de rostos, modificação do local, transformação da aparência), auditivos (substituição ou sobreposição de vozes, invenção de diálogos) ou audiovisuais preexistentes, de forma a fazer com que as pessoas acreditem na existência de algo que não ocorreu. Em outros casos, as falsidades profundas surgem da aplicação de ferramentas generativas, para a geração de registros fotográficos, áudios ou vídeos totalmente artificiais a partir de comandos específicos.

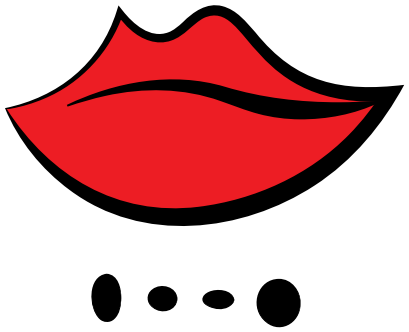
De forma resumida, as deepfakes são consideradas uma versão sofisticada de fake news. São, portanto, ferramentas de engano mais modernas e mais perigosas, capazes de imitar pessoas e simular acontecimentos reais, criando falsidades difíceis de serem detectadas.

QUAIS SÃO OS PRINCIPAIS TIPOS DE DEEFAKES?

Deepfakes podem surgir em diversas formas, a partir de técnicas variadas que possibilitam diferentes truques computacionais. As formas mais comuns de deepfakes envolvem:

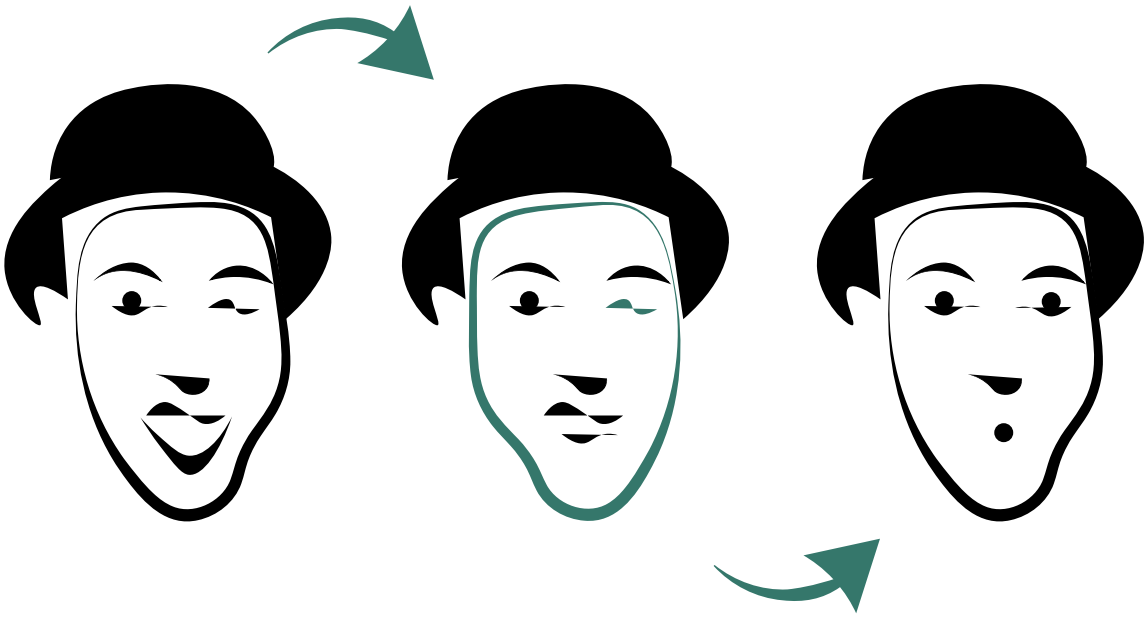
SUBSTITUIÇÃO
de um rosto por outro
(*face-swap*, ou troca-de-face)

CLONAGEM DE VOZ,
aparência e trejeitos, para gerar um vídeo em que a pessoa clonada reproduz todas as falas e movimentos realizados por um ator (pupper-master, ou jogo do ventríloquo).



Adulteração da região da boca, para que o movimento dos lábios acompanhe um áudio acrescentado (LIP-SYNC, ou sincronização labial)

As deepfakes mais sofisticadas são produzidas com o uso de redes generativas adversárias (também conhecidas como GANs, GENERATIVE ADVERSARIAL NETWORKS). Em uma GAN, dois algoritmos competem entre si: o primeiro com a função de gerar conteúdos falsos indetectáveis, e o segundo com a função de descobrir e apontar as falhas do primeiro. Dessa maneira, o primeiro algoritmo é constantemente aprimorado, conseguindo produzir resultados cada vez mais reais.

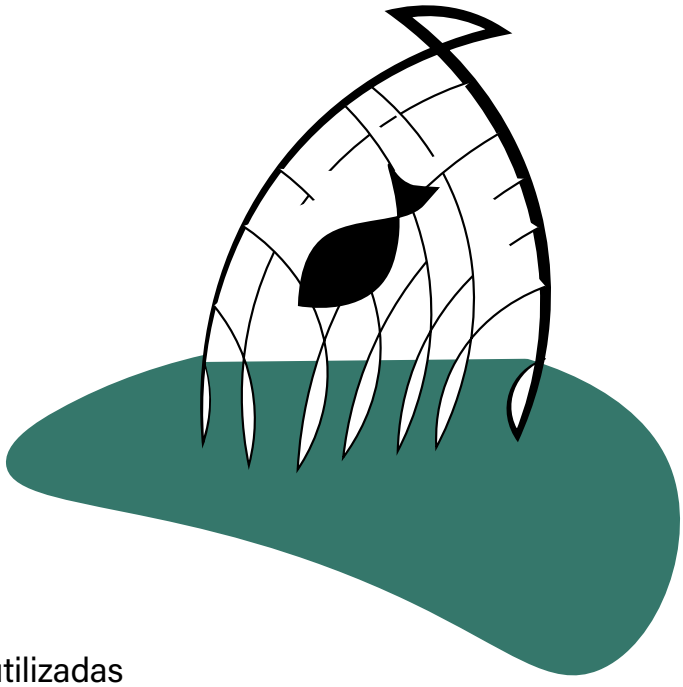


Em alguns casos, contudo, utilizam-se técnicas menos complicadas, como a simples redução da velocidade de fala para fazer parecer que o sujeito está bêbado. Nessas hipóteses, os vídeos adulterados recebem o nome de “cheapfakes”, por constituírem “falsificações baratas”, menos sofisticadas do que as deepfakes.

POR QUE É IMPORTANTE COMBATER AS DEEPFAKES?

As deepfakes podem nos fazer crer que alguém disse o que nunca diria, fez o que nunca faria ou esteve em alguma situação que jamais ocorreu. Várias situações podem decorrer disso, de golpes e fraudes diversas a tentativas de manipular a agenda pública e o debate democrático. Por exemplo, aqueles que as produzem podem nos enganar para obter informações pessoais sensíveis, como senhas bancárias e de contas de e-mail, ou mesmo para se fingir de parentes ou conhecidos para pedir empréstimos ou transferências financeiras.

Em algumas ocasiões, deepfakes já foram usadas para dar veracidade a falsos sequestros, com a exigência de pagamento de resgates. Já foram utilizadas, da mesma forma, para confundir funcionários de empresas, fazendo-os executar ordens de pagamento inexistentes, gerando a perda de grandes quantidades de dinheiro.



DANOS

As deepfakes também têm sido utilizadas como arma política, criando narrativas falsas que influenciam eleições no mundo todo. Têm viabilizado, do mesmo modo, conteúdos fraudulentos de teor impactante que reforçam percepções de intolerância, ódio e preconceito, e que por vezes terminam em episódios de violência física no mundo real.

USO MALÉFICO

Além disso, as deepfakes são utilizadas para gerar vídeos de teor sexual, inclusive em contextos de vingança ou assédio em ambiente escolar, a fim de expor pessoas e famílias a uma intensa dose de humilhação.

As deepfakes não fazem prova de nada. Pelo contrário, criam falsas representações de eventos não ocorridos, gerando graves prejuízos à sociedade, às organizações e às pessoas. Por isso, é importante saber identificá-las, recusá-las e combatê-las.

COMO IDENTIFICAR DEEPFAKES?

A depender do tipo e técnica utilizada, a identificação de deepfakes pode ser muito difícil, mesmo para especialistas. Algumas deepfakes são tão sofisticadas que o desmascaramento exige uma análise profissional, com o apoio de técnicas e ferramentas de perícia.

DEEPFAKE DE VÍDEOS

De todo modo, há algumas falhas comuns que, se bem observadas, denunciam a existência de um conteúdo fraudulento.

Em deepfakes de vídeos, estes são os defeitos mais comuns:

1. Diferenças entre o tom da pele do corpo e a tonalidade do rosto
2. Anormalidades na movimentação dos olhos ou falta de naturalidade nas expressões
3. Ausência de marca da pessoa, como uma pinta, tatuagem ou sinal
4. Alterações, ainda que leves, no sotaque, na entonação ou na voz
5. Falta de sincronia entre o movimento dos lábios e a fala
6. Rigidez ou falta de naturalidade no movimento corporal
7. Elementos com aparência artificial, como o cenário, a iluminação, as cores e a própria expressão das pessoas reproduzidas no vídeo.

DEEPFAKE DE IMAGENS

Saiba identificar os indícios:

FUNDO
ou paisagem não condiz com o local dos fatos

SOMBRA
apagada ao redor dos olhos

TONALIDADE
de cor diferente com as demais partes do corpo

ASSIMETRIA
vísivel: ausência de brinco em uma das orelhas

DETALHES
estranhos, como rosto, pernas e braços desproporcionais

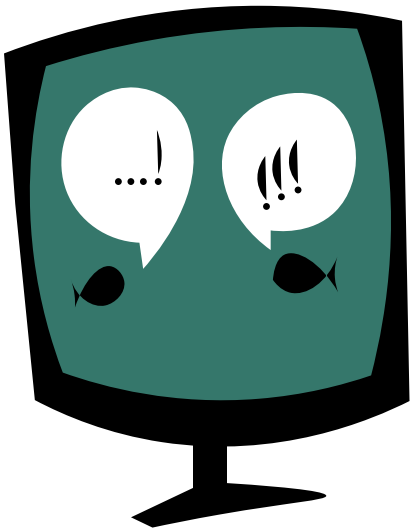
ENCAIXE
imperfeito entre a cabeça e o pescoço e algo estranho no interior da boca, pois em geral as ferramentas não são boas em reproduzir a língua

ROUPAS
usadas pelos envolvidos são incompatíveis com as características do lugar e das demais pessoas que aparecem no registro

Em caso de dúvida, reveja o conteúdo em tela cheia, preferencialmente em telas grandes, como em tablets ou computadores. A ampliação facilita a visualização de falhas na produção de deepfakes. Teste suas habilidades de detecção de imagens falsas acessando o projeto da Universidade de Northwestern: https://detectfakes.kellogg.northwestern.edu/

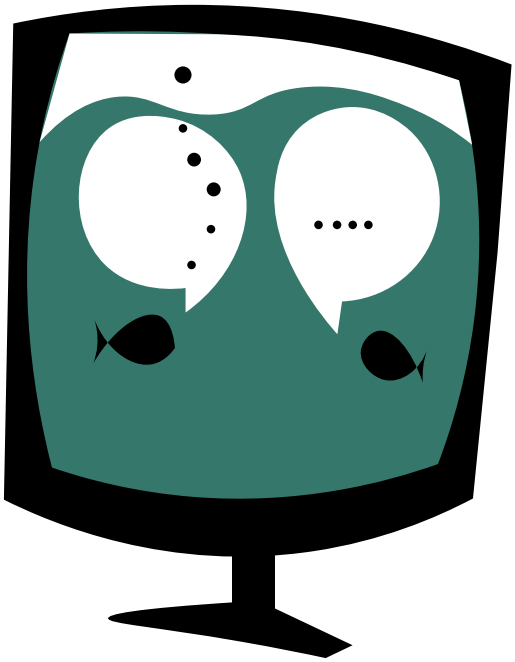
DEEPFAKE DE ÁUDIO

Em deepfakes de áudio, confira se:



1 VOZ, SOTAQUE, ENTONAÇÃO E VOCABULÁRIO são totalmente compatíveis com os registros vocais e com o modo de ser da pessoa em questão

2 Em diálogos, há consistência no TEMPO de reação entre as falas dos interlocutores



3 Há diferenças de VOLUME ou equalização entre as falas de cada participante

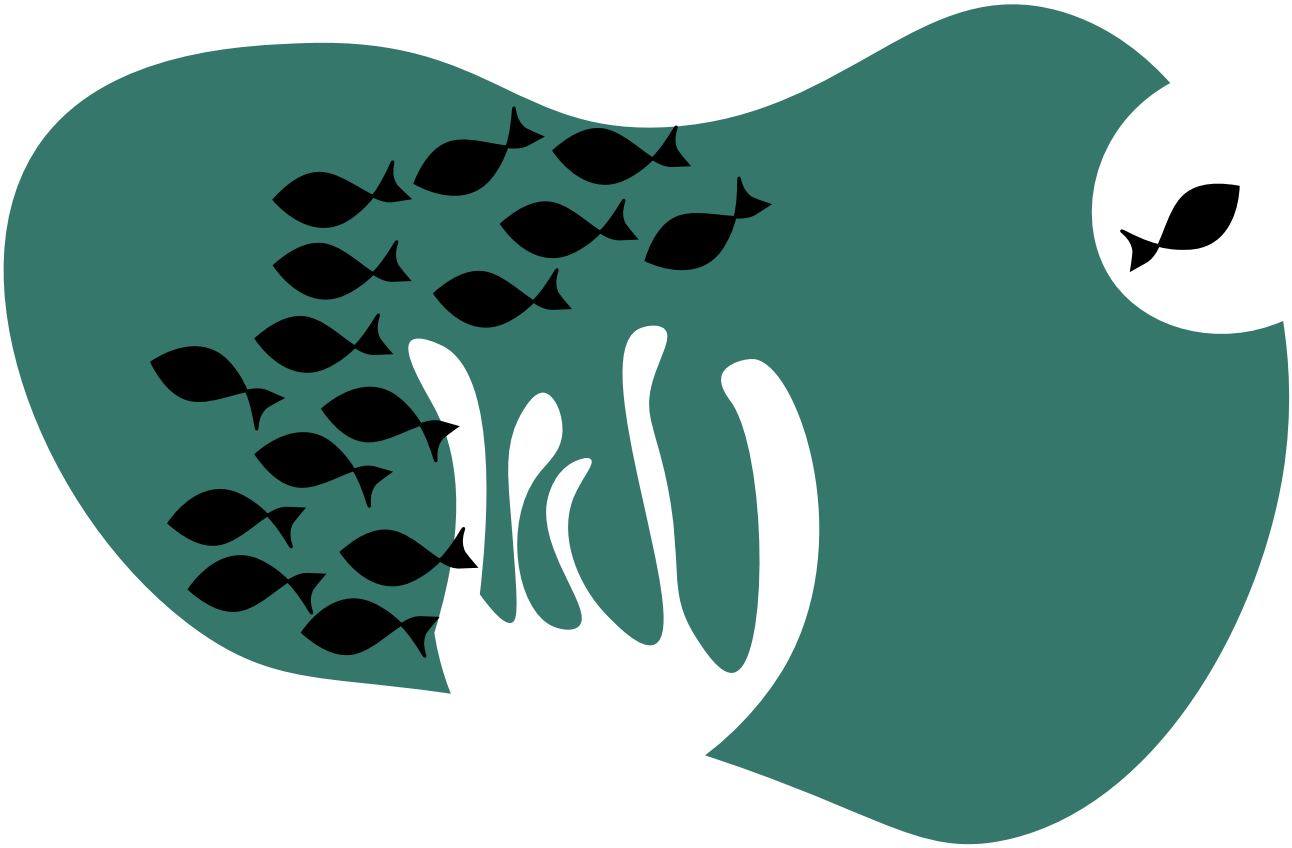
4 Há INTERRUPÇÕES abruptas de palavras, frases ou ideias, denunciando a presença de cortes ou edições

Atente-se sempre para a origem das mensagens, questionando se as fontes que as publicaram têm algo a ganhar com a sua divulgação. Análises contextuais também ajudam, tanto quando o conteúdo for sobre alguém que você conhece como sobre uma pessoa pública. A situação representada parece exagerada ou até absurda, "fora do personagem"? Procure se atentar ainda mais para as dicas acima nesses casos, pois pode ser uma deepfake.

Além disso, tratando-se de questões sérias, impactantes ou polêmicas, é importante conferir se a veracidade do conteúdo foi confirmada por alguma agência de checagem de fatos ou por um veículo confiável de comunicação.

COMO DENUNCIAR DEEPFAKES?

Em redes sociais e plataformas de vídeo, as falsidades profundas podem ser denunciadas por meio dos canais destinados ao envio de apontamentos sobre violação de regras de comunidade ou termos de uso.



CONTEXTO ELEITORAL

Se as falsidades atentam contra a integridade das eleições, como no caso de deepfakes contra a urna eletrônica ou o trabalho realizado pela Justiça Eleitoral, **as denúncias podem ser feitas no seguinte endereço:**

https://www.tse.jus.br/eleicoes/sistema-de-alertas

Notícias sobre a desinformação sobre o processo eleitoral em redes sociais também podem ser encaminhadas ao Tribunal Superior Eleitoral por telefone, por meio do disque-denúncia SOS Voto, através do número 1491.

SOS VOTO
1491

O SOS Voto funciona de segunda a sexta, das 8h às 20h, e no sábado das 9h às 17h, tendo capacidade para receber até mil ligações diárias. Por fim, se as deepfakes envolverem a possibilidade de crime, como usurpação de identidade, golpes ou atentados contra a honra ou a imagem das pessoas, é possível reportá-las pelo telefone, acessando o Disque Denúncia (181).

DISQUE DENÚNCIA
181

SOBRE O PROGRAMA DE COMBATE À DESINFORMAÇÃO

O Programa de Combate à Desinformação foi instituído pela Resolução nº 742, de 27 de agosto de 2021, em harmonia com o sistema de proteção das liberdades de comunicação, previsto na Constituição Federal de 1988, e com a Convenção Americana sobre Direitos Humanos. A criação do programa está inserida no contexto do Objetivo de Desenvolvimento Sustentável nº 16 da Agenda 2030 da Organização das Nações Unidas (Paz, Justiça e Instituições Eficazes), à qual o Supremo aderiu integralmente. Desde outubro de 2023, o Programa é conduzido pela Coordenadoria de Combate à Desinformação (CCOD), e executado em colaboração com outras áreas de atuação do STF.

Acesse: Supremo Tribunal Federal (stf.jus.br)

SOBRE A DATA PRIVACY BRASIL

A Data Privacy Brasil é uma organização que nasce da parceria entre uma escola e uma associação de pesquisa com o objetivo de fomentar a cultura de proteção de dados e direitos digitais no Brasil e no mundo, promover direitos fundamentais e valores ligados à justiça social diante das tecnologias contemporâneas e processos de datificação. Por meio da educação, da pesquisa, da sensibilização e da mobilização da sociedade, buscamos uma sociedade democrática na qual as tecnologias estejam a serviço da autonomia, dignidade das pessoas e redução de assimetrias de poder.

Acesse: https://www.dataprivacybr.org/

SOBRE AS ILUSTRAÇÕES

"A ilusão, a manipulação da realidade e a encenação das deepfakes são retratadas de modo leve pela alegoria de personagens cênicos que escondem seus rostos, mas revelam técnicas para desmascarar as deepfakes. Em sintonia, a metáfora dos cardumes e seus elementos representam o comportamento social no ambiente digital por meio das possíveis armadilhas do mundo marinho."

Por Renato Barros de Carvalho


STF
SUPREMO TRIBUNAL FEDERAL


DataPrivacyBR
Research


PROGRAMA DE COMBATE À
DESINFORMAÇÃO
SOCIEDADE INFORMADA, DEMOCRACIA FORTE