

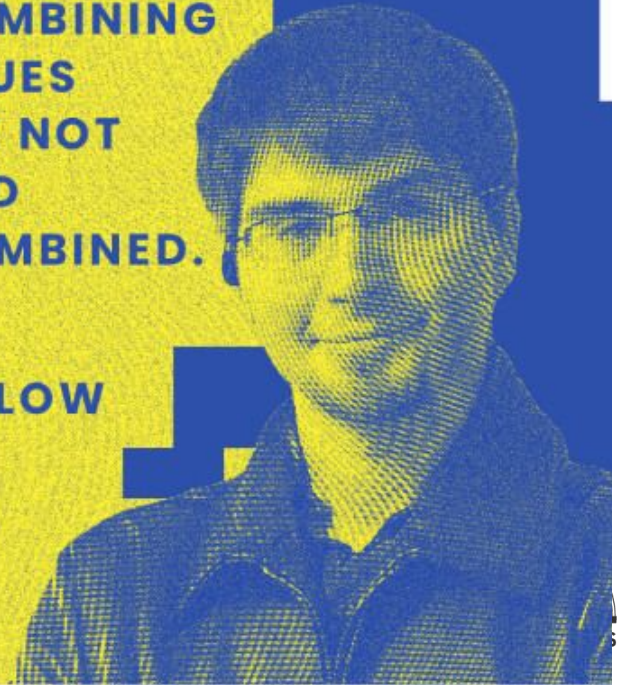
Aplicação e implicações da IA nas eleições e na disseminação de informações, com a finalidade de debater sobre os desafios e os riscos que a IA apresenta à integridade jornalística e democrática

Diretora de Projeto - ITS Rio
celina@itsrio.org

"In the future, people should no longer believe images and videos on the internet as a matter of course".

**THE MOST
INTERESTING
RESULTS COME
FROM COMBINING
TECHNIQUES
THAT ARE NOT
SUPPOSED
TO BE COMBINED.**

**IAN
GOODFELLOW**



1. Evolução do uso de *deepfakes*
2. Uso de *deepfakes* em contexto de guerra e eleição
3. Iniciativas para lidar com esse desafio



[Submitted on 10 Jun 2014]

Generative Adversarial Networks

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio

We propose a new framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G . The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game. In the space of arbitrary functions G and D , a unique solution exists, with G recovering the training data distribution and D equal to $1/2$ everywhere. In the case where G and D are defined by multilayer perceptrons, the entire system can be trained with backpropagation. There is no need for any Markov chains or unrolled approximate inference networks during either training or generation of samples. Experiments demonstrate the potential of the framework through qualitative and quantitative evaluation of the generated samples.

Subjects: **Machine Learning (stat.ML)**; Machine Learning (cs.LG)

Cite as: [arXiv:1406.2661](https://arxiv.org/abs/1406.2661) [stat.ML]

(or [arXiv:1406.2661v1](https://arxiv.org/abs/1406.2661v1) [stat.ML] for this version)

<https://doi.org/10.48550/arXiv.1406.2661> 

Submission history

From: Ian Goodfellow [[view email](#)]

[v1] Tue, 10 Jun 2014 18:58:17 UTC (1,257 KB)

2015



2015 - tecnologia ainda estava sendo aprimorada. Ainda não era possível criar imagens realistas. Alteração era facilmente identificada.



Instituto
de Tecnologia
& Sociedade
10 anos



2016



2016 - Já era possível manipular imagem colocando óculos, por exemplo.



Instituto
de Tecnologia
& Sociedade
10 anos



2017



The faces generated in 2017 surpass previous results, and some are barely recognizable as AI products. Image: Nvidia 2017

2017 - Nvidia - Salto de qualidade. Começava com imagem com resolução baixa e gradativamente a resolução vai melhorando.



Cars



Bedrooms

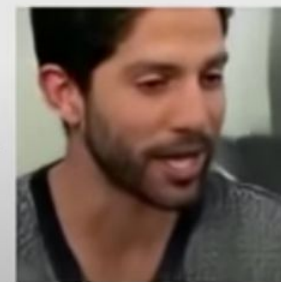
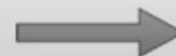
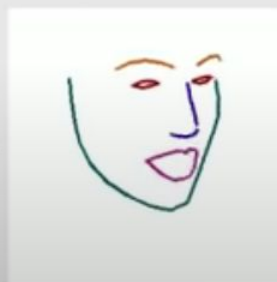
Tecnologia usada não só usando rostos, mas também objetos.

Learning talking heads from few examples

Training frames:



Landmarks
detector



Driving sequence

Face landmarks

Learned talking head

2019 - evolução

2019 - Deepfakes começaram a ser usados em sites de pornografia.

Canais de youtube especializados em deepfakes

Scarlett Johansson: Fight Against 'Deepfake' Porn Is a 'Lost Cause'

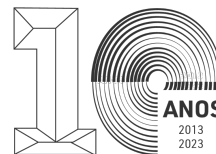
"Nothing can stop someone from cutting and pasting my image or anyone else's onto a different body," star tells The Washington Post



Usos de Deepfakes no contexto eleitoral e de guerra



Instituto
de Tecnologia
& Sociedade
10 anos





Em 2018, o comediante Jordan Peele criou um vídeo no qual o Barack Obama insultava o então candidato Donald Trump.

[MARKETS](#)[BUSINESS](#)[INVESTING](#)[TECH](#)[POLITICS](#)[CNBC TV](#)[INVESTING CLUB](#)[PRO](#)[NEW!](#)

PIMCO Investments LLC, distributor. ©2023, PIM

TECH

Fake videos could be the next big problem in the 2020 elections

PUBLISHED TUE, OCT 15 2019•1:26 AM EDT | UPDATED FRI, JAN 17 2020•2:49 AM EST



Grace Shao
[@GRACEMZSHAO](#)

SHARE



KEY POINTS

- Fake news was a big problem for the 2016 election. “Deepfake” videos could be an even bigger problem in 2020.
- States including California and Texas have enacted laws that make deepfakes illegal when they’re used to interfere with elections.

VALL
CON

Manipulated images

**The video falsely shows
the Ukrainian President
telling his forces...**

No início da guerra entre Ucrânia e Rússia surgiu um video do presidente Zelensky pedindo para o povo se entregar e baixar as armas.

O que fazer?



Instituto
de Tecnologia
& Sociedade
10 anos



Iniciativas legislativas para enfrentar o problema:

EUA:

- California e Texas aprovaram leis proibindo o uso de deepfakes em contexto eleitoral.
- "DeepFake Accountability Act": Projeto de Lei Federal que criminaliza a produção de deepfakes que não tenha um identificador/marca d'água indicando que o conteúdo foi gerado artificialmente e de quem tenta retirar essa identificação.

China:

- Lei contra Deepfake - todo conteúdo gerado artificialmente tem que ser rotulado com tal, deve haver o consentimento da pessoa antes de ter sua imagem ou voz alterada além de medidas de informar ao governo sobre a criação de deepfakes.
- Governo criou uma lista de 41 empresas que estariam de acordo com os requisitos da nova legislação.

Canadá:

- Prevenção: campanha de conscientização.
- Detecção: Governo está investindo em tecnologia para aperfeiçoar formas de identificar deepfake.
- Resposta: Legislação para criminalizar a produção de deepfake com intenção de causar confusão e dano.

União Europeia:

- DSA - medidas para lidar com deepfakes - multa de 6%
- AI Act - art 52 -3 - obrigação de informar que conteúdo foi artificialmente gerado.

Coréia do Sul:

- Torna ilegal a distribuição de deepfakes com intuito de "causar dano ao interesse público".

Brasil

- PL 1002/2023 - altera lei eleitoral para criminalizar deepfake
- Art. 57-H. § 3º Constitui crime criar, utilizar e propagar deep fake no período das eleições visando manipular informações para prejudicar o senso de realidade dos eleitores com o intuito de desconstruir reputações políticas, punível com detenção de 2 (dois) a 4 (quatro) anos e multa de R\$ 15.000,00 (quinze mil reais) a R\$ 50.000,00 (cinquenta mil reais). I - Para fins desta Lei considera-se deep fake a criação de vídeos a partir de inteligência artificial que reproduzem a aparência, as expressões e a voz de algum candidato.”
- PL 2338/22 - Art. 7º - obrigações de transparência

Outras Iniciativas:

Taiwan:

- Criando um 'nerd immunity' - treinando centenas de checadores de fato e treinando as pessoas para identificar conteúdo que foi gerado artificialmente.

Japão:

- "Perfil Originador" (Originator profile) - Consórcio entre universidades e empresas de mídia para fornecer informações sobre criadores de conteúdo da web, operadores de sites, anunciantes, apontando a origem e credibilidade da fonte.

Article



Japan Meteorological Agency's volcanic alerts system found lacking

Publishing date Aug. 10, 2022
Last Updated July 3, 2022 9:00
Editor Taro Yamada
Writer Digital editors
[more details about this article ...](#)

ABCD
Publishing

Publishing organization
ABCD PUBLISHING

 **CERTIFIED**

**against Invalid Traffic and other 2 certification valid thru Dec, 2022**
✓ valid period

Publish Guideline
[Principles](#) 

Outras Iniciativas da Indústria:

Content Authenticity Initiative

- Plataforma criada pela indústria para garantir a originalidade do conteúdo produzido. Sistema que detecta a proveniência de conteúdo digital por meio de desenvolvimento de código aberto, colaboração entre setores e interoperabilidade de ferramentas.

Facebook, Youtube

- Baniram o uso de deepfakes em suas plataformas desde que não sejam claramente identificadas. Exceção para paródias/humor.

Intel

- Fake Catcher - Primeiro detector de deepfake em tempo real do mundo e com taxa de precisão de 96%.

Lembrando que nem todo uso
de *deepfake* é negativo...





Imagens criadas artificialmente de sobreviventes do holocausto para contar suas memórias.

Lyrebird AI

Using artificial intelligence to enable creative expression.

Ajuda os pacientes que perderam a habilidade de falar e clonando a voz, eles conseguem "voltar a falar"



BOSTON, USA
(BEFORE)



HOMS, SYRIA
(PRESENT)



BOSTON, USA
(AFTER)



Aumentando empatia em relação a vítimas de desastres.

Resumindo...

- Não tem solução única
- Nem todo uso de Deepfake é negativo
- Lei proibindo o uso não resolve, tecnologia evolui e criam métodos para remoção de identificadores.
- Lei eleitoral no Brasil já proíbe "fatos sabidamente inverídicos em relação a partidos políticos ou a candidatos, capazes de exercer influência perante o eleitorado."
- Campanha de educação midiática, ensinando olhar crítico sobre conteúdo online.

Referências

<https://www.responsible.ai/post/a-look-at-global-deepfake-regulation-approaches>

<https://www.canada.ca/en/democratic-institutions/news/2019/01/government-of-canada-unveils-plan-to-safeguard-canadas-election.html>

<https://www.descript.com/lyrebird>

[https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU\(2021\)690039_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf)

<https://the-decoder.com/history-of-deepfakes/>

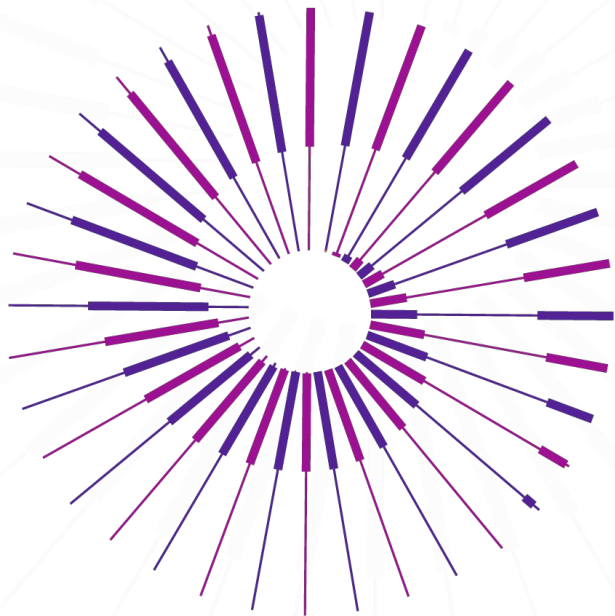
<https://www.eff.org/deeplinks/2019/06/congress-should-not-rush-regulate-deepfakes>

<https://theconversation.com/how-ai-could-take-over-elections-and-undermine-democracy-206051>

HÁ 10 ANOS

PELA INTERNET

ABERTA



Agradeço pela atenção!

Até a próxima,

ITS Rio

www.itsrio.org

Celina Bottino

celina@itsrio.org

contato@itsrio.org

@itsriodejaneiro



Instituto
de Tecnologia
& Sociedade
10 anos

