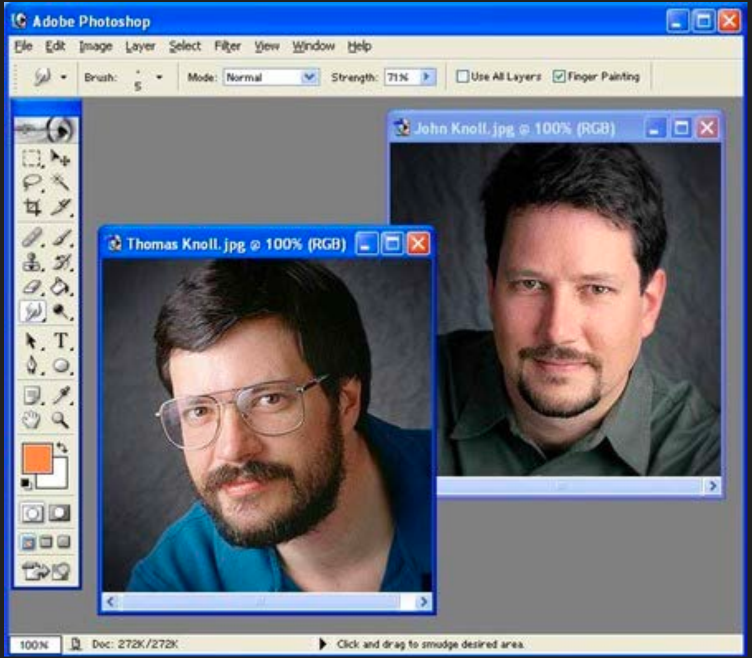
Detecting, Combating, and Identifying Dis- and Mis-information

Chris Bregler, Google Research

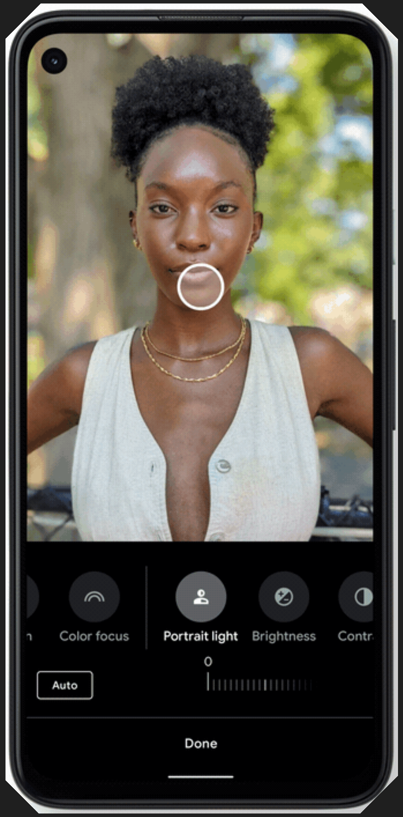
Democratization of Media Manipulation



Brush: 5 Mode: Normal Strength: 71% □ Use All Layers ☑ Finger Painting

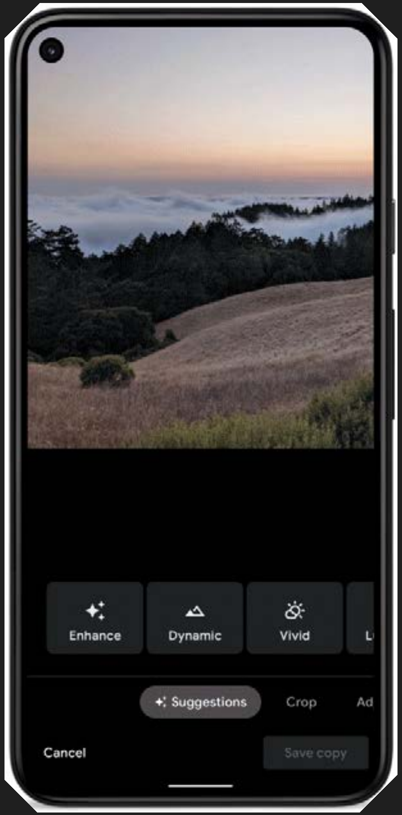
100% Doc: 272K/272K Click and drag to smudge desired area.

Portrait Lighting (Pixel)



Google Research/DayDream + UCSD:
Sun, Barron, Tsai, Xu, Yu, Fyffe,
Rhemann, Busch, Debevec,
Ramamoorthi + Google Product Team

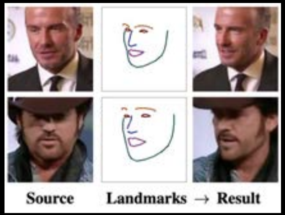
HDR & Sky Palette Transfer



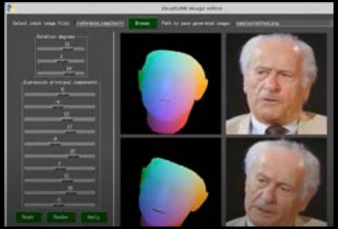
Google Research / Pixel / Photos:
Pritch, Dorado, Balke, Liba, Talebi,
Kelly, Sarma, Milne, Cai, Chen,
Manoogian, Maffeo, Meron, Wang,
Howard, Eben, Kanazawa,
Movshovitz-Attias, Milanfar,
Campbell, Ma, Ng, Khattar,
Bleibel,Hasinoff

Academic Research / Companies

Thies et al 2016



Zakharov et al 2019



Doukas et al 2021

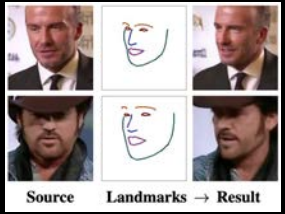
Synthesia

MyHeritage

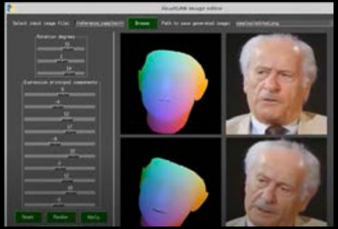
NVIDIA 2021

Academic Research / Companies

Thies et al 2016



Zakharov et al 2019



Doukas et al 2021

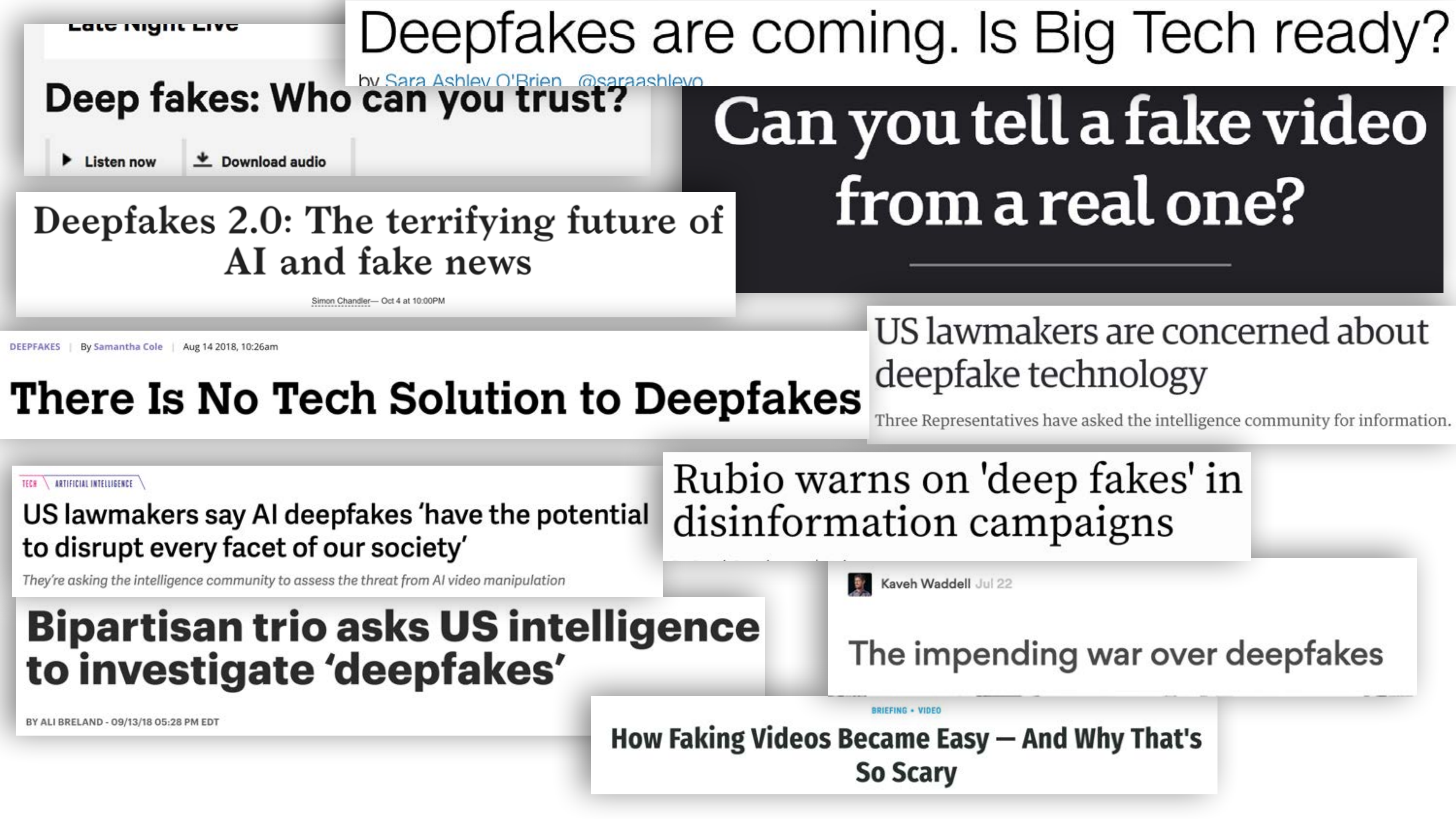
Synthesia

MyHeritage

NVIDIA 2021

DeepFakes

- Discussed on reddit by anonymous user “/u/deepfakes” in fall 2017
- First public code: Dec-15-2017
- FakeApp in Spring 2018
- Every few days new code



Late Night Live

Deep fakes: Who can you trust?

▶ Listen now

⤓ Download audio

Deepfakes are coming. Is Big Tech ready?

by Sara Ashley O'Brien @saraashleyo

Can you tell a fake video from a real one?

Deepfakes 2.0: The terrifying future of AI and fake news

Simon Chandler— Oct 4 at 10:00PM

DEEPFAKES | By Samantha Cole | Aug 14 2018, 10:26am

There Is No Tech Solution to Deepfakes

US lawmakers are concerned about deepfake technology

Three Representatives have asked the intelligence community for information.

TECH \ ARTIFICIAL INTELLIGENCE \

US lawmakers say AI deepfakes 'have the potential to disrupt every facet of our society'

They're asking the intelligence community to assess the threat from AI video manipulation

Rubio warns on 'deep fakes' in disinformation campaigns

Bipartisan trio asks US intelligence to investigate 'deepfakes'

BY ALI BRELAND - 09/13/18 05:28 PM EDT

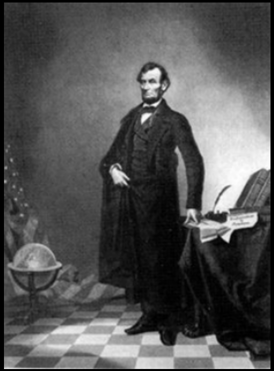
Kaveh Waddell Jul 22

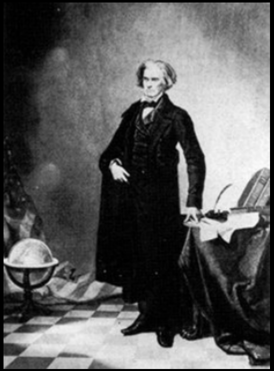
The impending war over deepfakes

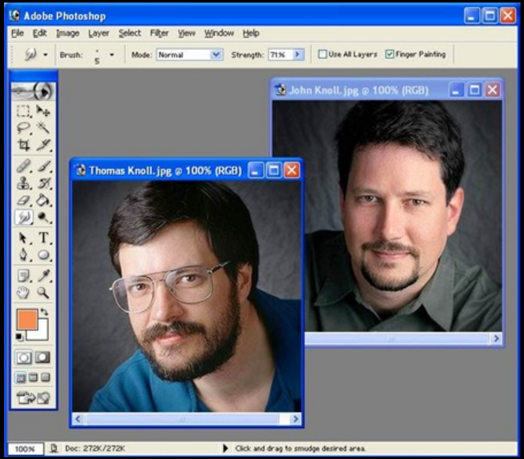
BRIEFING • VIDEO

How Faking Videos Became Easy — And Why That's So Scary

Manipulated Media is as old as Invention of Camera







Adobe Photoshop

Brush: 5

Mode: Normal

Strength: 71%

□ Use All Layers

☑ Finger Painting

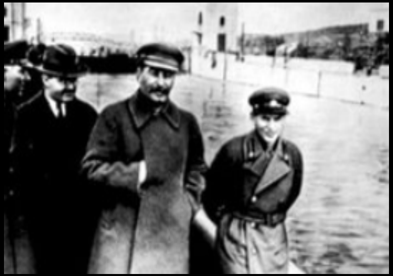
John Knoll.jpg @ 100% (RGB)

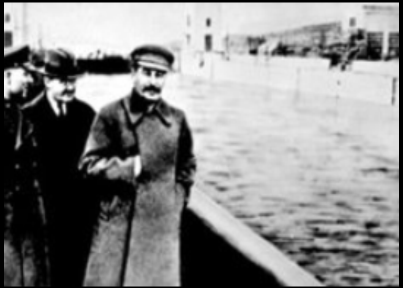
Thomas Knoll.jpg @ 100% (RGB)

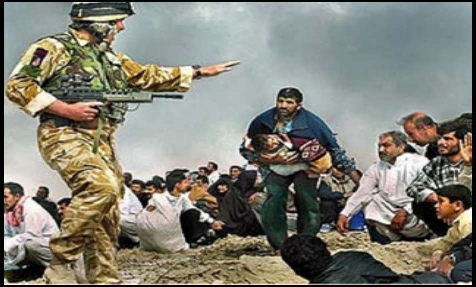
100%

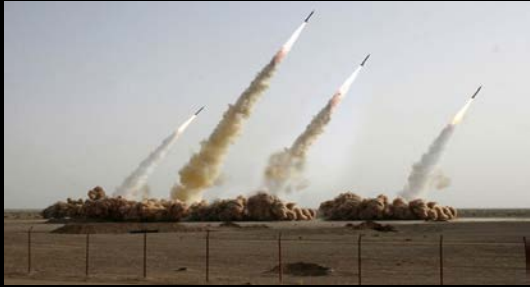
Doc: 272K/272K

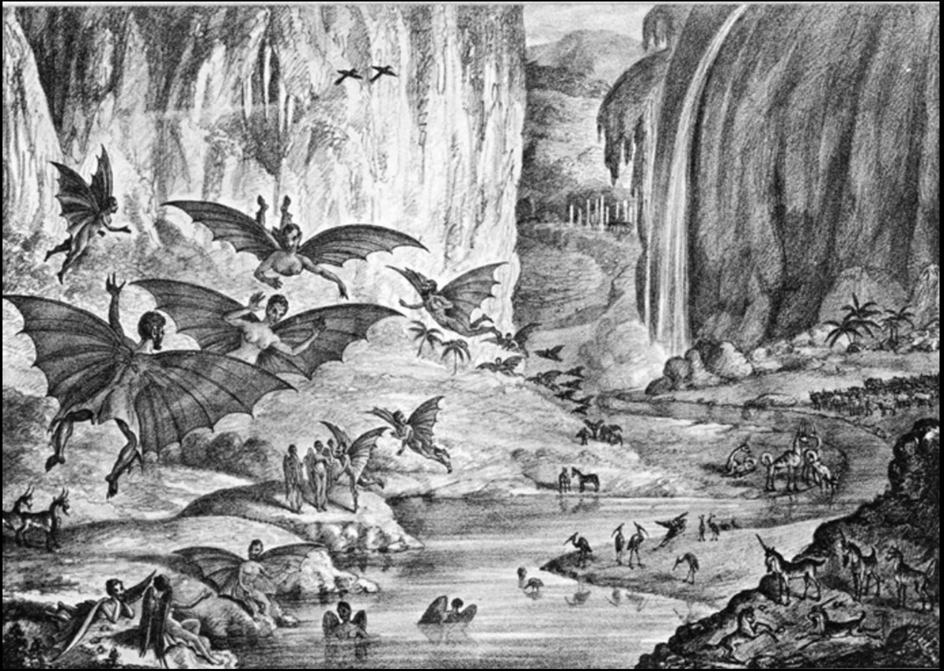
Click and drag to smudge desired area.











The open web empowers anyone anywhere to be a source of information

Google News Initiative: Address this challenge in three ways

Google

We build products to elevate quality journalism on our platforms

YouTube

Newsrooms

We collaborate with newsrooms to surface accurate information

CREDIBILITY COALITION

The Trust Project

FIRST DRAFT

W3C Word Wide Web Consortium

Audiences

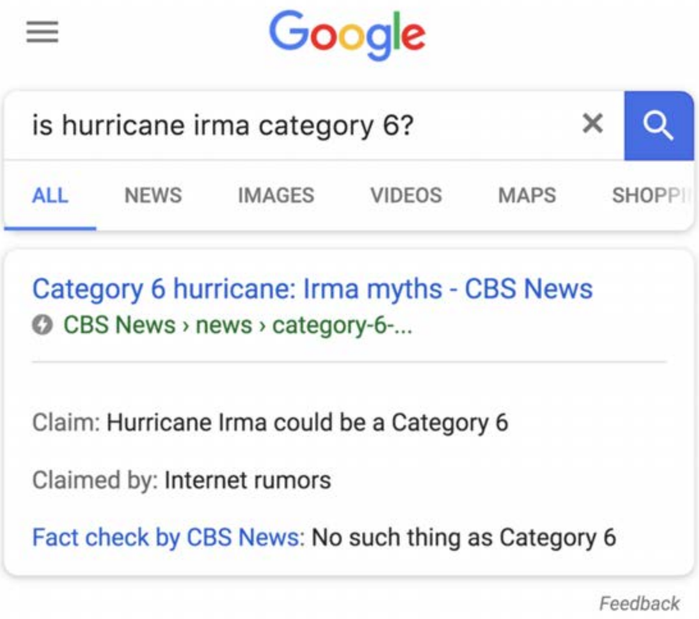
We support research and build programs to improve digital literacy

The Canadian Journalism Foundation

Stanford University

theguardian foundation

Verified fact checkers throughout the ecosystem



is hurricane irma category 6?

Category 6 hurricane: Irma myths - CBS News

CBS News › news › category-6-...

Claim: Hurricane Irma could be a Category 6

Claimed by: Internet rumors

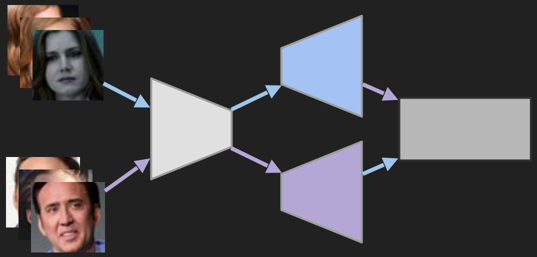
Fact check by CBS News: No such thing as Category 6

IFCN
@Poynter.

Back to Visual Misinfo / Manipulated Media

Internal Efforts: DeepFakes, CheapFakes

Synthesis:

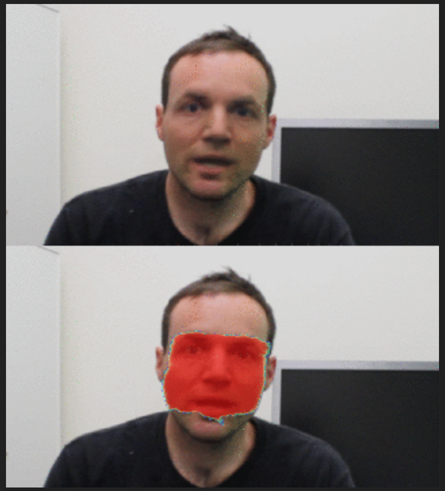


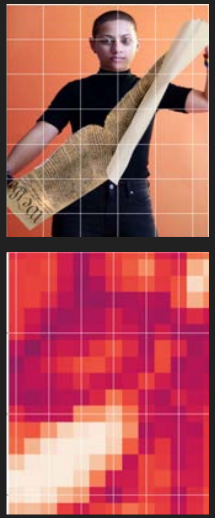
Pre/Post-2018: internal + opensource

Long History (GAN, VAE, VFX, ...)

Datasets:

Detectors:





Dufour, Leung, Sud et al. with Verdoliva

Google’s DeepFake Detection Data release V1 (with JigSaw)

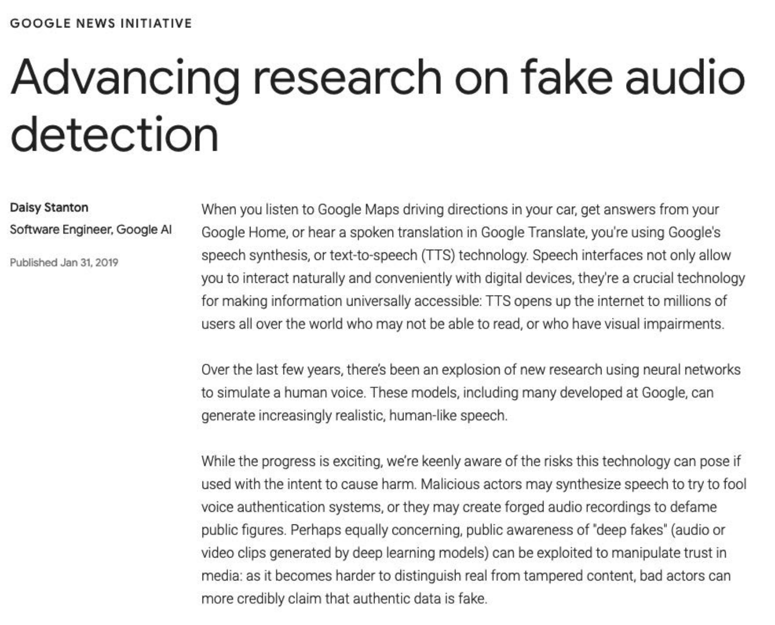
Technische Universität München

Università degli Studi di Napoli Federico II

By: Nick Dufour, Andrew Gully, Per Karlsson, Alexey Victor Vorobyov, Thomas Leung, Jeremiah "Spudde" Childs, Christoph Bregler,
With: Andreas Roessler, Davide Cozzolino, Justus Thies, Luisa Verdoliva, Matthias Niessner.

Google’s DeepFake Detection Data release V1

Audio DeepFakes



GOOGLE NEWS INITIATIVE

Advancing research on fake audio detection

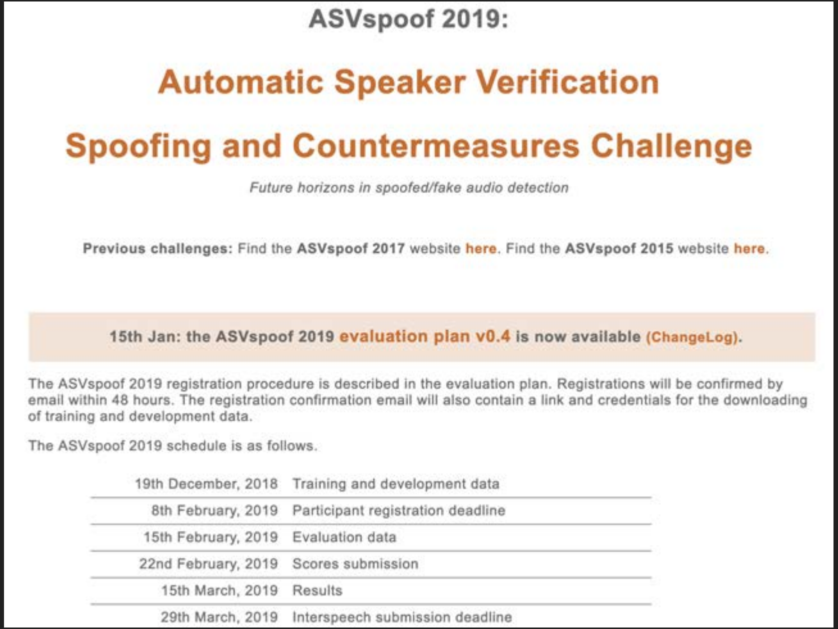
Daisy Stanton
Software Engineer, Google AI

Published Jan 31, 2019

When you listen to Google Maps driving directions in your car, get answers from your Google Home, or hear a spoken translation in Google Translate, you're using Google's speech synthesis, or text-to-speech (TTS) technology. Speech interfaces not only allow you to interact naturally and conveniently with digital devices, they're a crucial technology for making information universally accessible: TTS opens up the internet to millions of users all over the world who may not be able to read, or who have visual impairments.

Over the last few years, there's been an explosion of new research using neural networks to simulate a human voice. These models, including many developed at Google, can generate increasingly realistic, human-like speech.

While the progress is exciting, we're keenly aware of the risks this technology can pose if used with the intent to cause harm. Malicious actors may synthesize speech to try to fool voice authentication systems, or they may create forged audio recordings to defame public figures. Perhaps equally concerning, public awareness of "deep fakes" (audio or video clips generated by deep learning models) can be exploited to manipulate trust in media: as it becomes harder to distinguish real from tampered content, bad actors can more credibly claim that authentic data is fake.



ASVspoof 2019:

Automatic Speaker Verification

Spoofing and Countermeasures Challenge

Future horizons in spoofed/fake audio detection

Previous challenges: Find the **ASVspoof 2017** website **here**. Find the **ASVspoof 2015** website **here**.

15th Jan: the ASVspoof 2019 evaluation plan v0.4 is now available (ChangeLog).

The ASVspoof 2019 registration procedure is described in the evaluation plan. Registrations will be confirmed by email within 48 hours. The registration confirmation email will also contain a link and credentials for the downloading of training and development data.

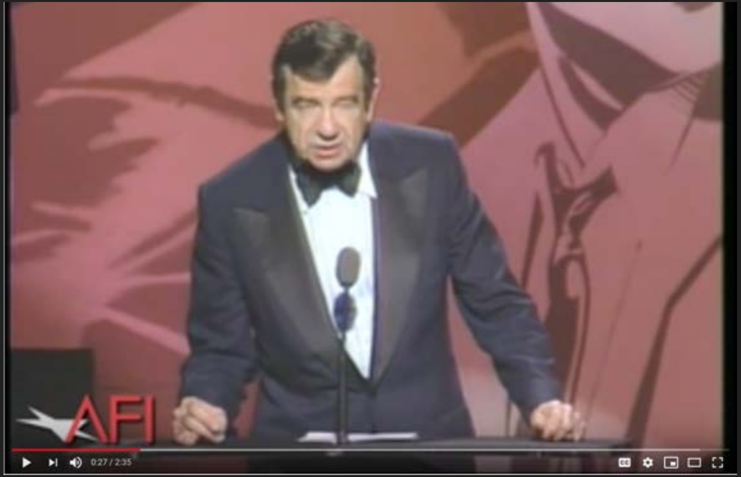
The ASVspoof 2019 schedule is as follows.

19th December, 2018	Training and development data
8th February, 2019	Participant registration deadline
15th February, 2019	Evaluation data
22nd February, 2019	Scores submission
15th March, 2019	Results
29th March, 2019	Interspeech submission deadline

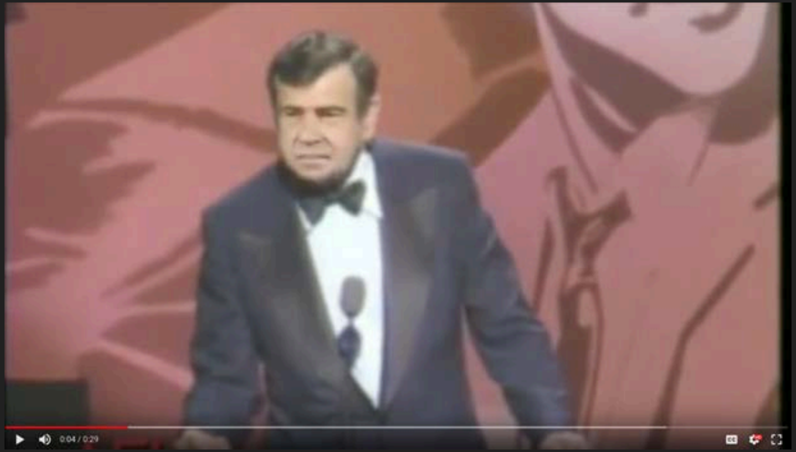
Informing Users: Jigsaw + Our Research Team

Open question: Even given an oracle, would it be useful?

Jigsaw conducted a UX study using deepfakes generated by our team:



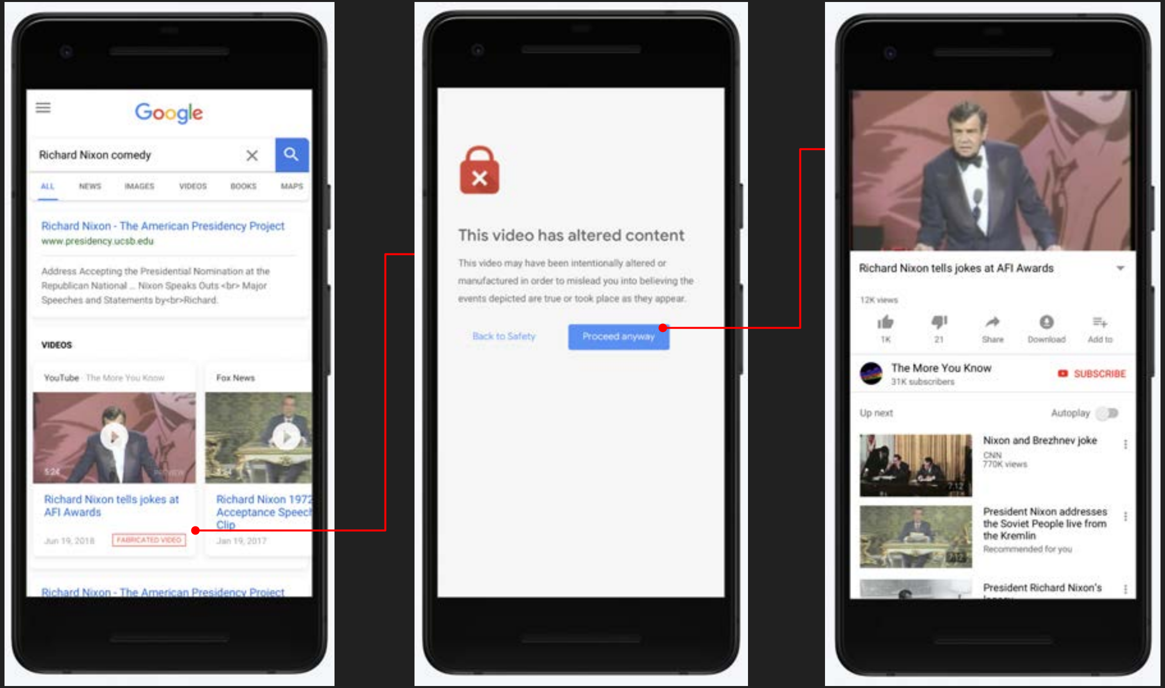
Walter Matthau telling jokes at AFI Awards (real)



Richard Nixon telling jokes at AFI Awards (fake)

Informing Users: Jigsaw + Our Research Team

Jigsaw created mockup UI elements and assessed their efficacy (*N*=820)



Richard Nixon - The American Presidency Project
www.presidency.ucsb.edu

Address Accepting the Presidential Nomination at the Republican National ... Nixon Speaks Outs
 Major Speeches and Statements by
Richard.

VIDEOS

YouTube · The More You Know
Richard Nixon tells jokes at AFI Awards
Jun 19, 2018 FABRICATED VIDEO

Fox News
Richard Nixon 1972 Acceptance Speech Clip
Jan 19, 2017

Richard Nixon - The American Presidency Project

This video has altered content

This video may have been intentionally altered or manufactured in order to mislead you into believing the events depicted are true or took place as they appear.

Richard Nixon tells jokes at AFI Awards
12K views

The More You Know
31K subscribers

Nixon and Brezhnev joke
CNN
770K views

President Nixon addresses the Soviet People live from the Kremlin
Recommended for you

President Richard Nixon's

https://link.medium.com/AnVgvBetXjb

© Bloomberg via Getty Images

[1] Chesney & Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy and National Security"

Liar's Dividend

- Joao Doria, former Sao Paulo mayor, seen in a sex tape 5 days before election.
- He claims it's a deepfake, we (and others) think it's probably authentic.
- Electorate split on whether tape is authentic; Doria wins reelection.
- Example of the Liar's Dividend: "...in what might be understood as a 'liar's dividend,' deep fakes make it easier for liars to avoid accountability for things that are in fact true."[1]

© Bloomberg via Getty Images

[1] Chesney & Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy and National Security”

JigSaw/Google Assembler: Detectors -> Fact Checkers

UNIVERSITY OF MARYLAND

Naples

Berkeley UNIVERSITY OF CALIFORNIA

(Dartmouth)

JIGSAW

Google AI

ANIMAL POLITICO

Africa Check Sorting fact from fiction

RAPPLER

AFP

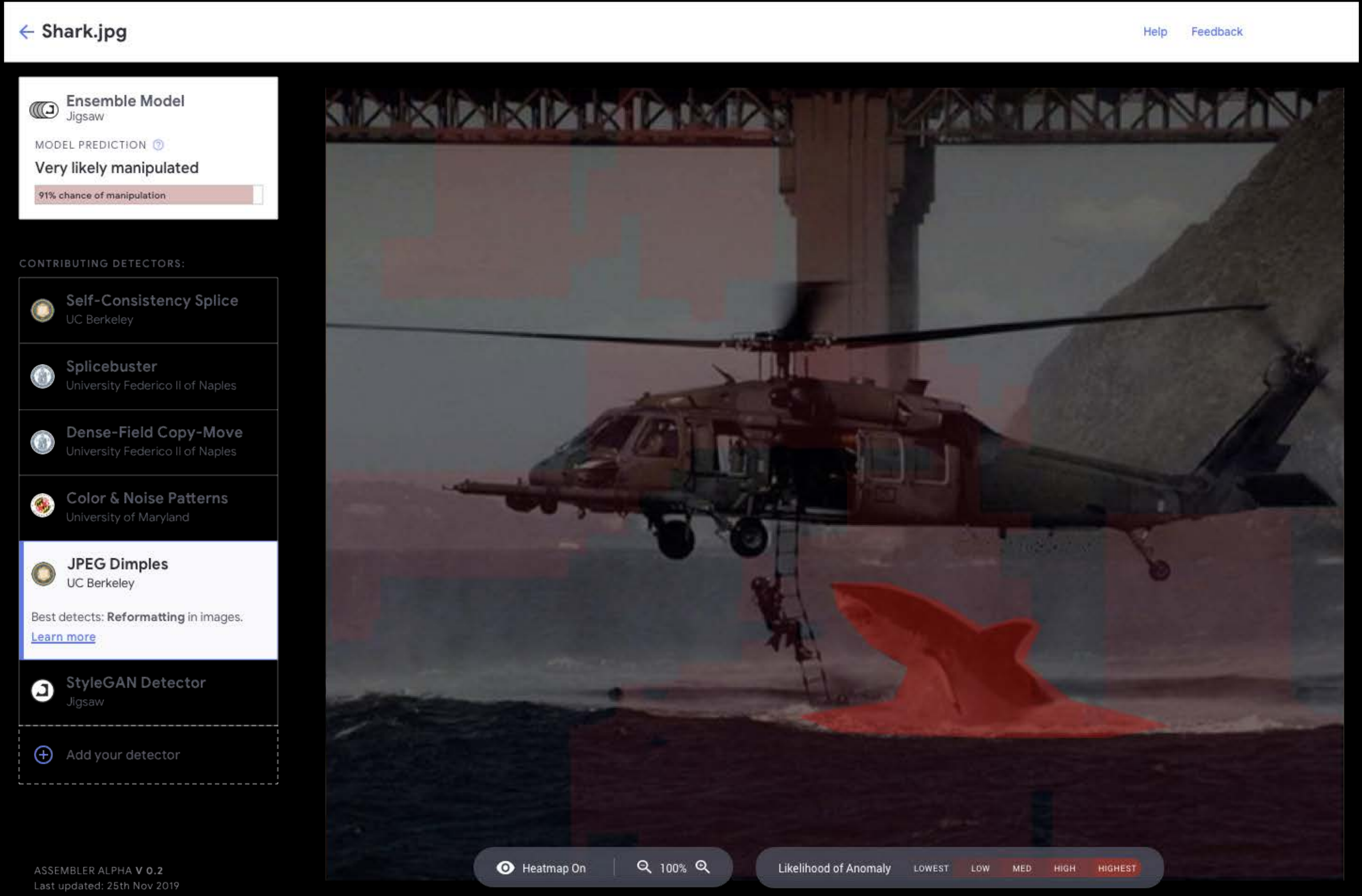
CODE for AFRICA

Le Monde.fr

Current Academics: Davide Cozzolino, Giovanni Poggi, Luisa Verdoliva, Minyoung Huh, Andrew Liu, Andrew Owens, Ayosha Efros, Shruti Agarwal, Hany Farid, Peng Zhou, Xinton Han, Vlad Morariu, Abhinav Shrivastava, Larry Davis

Jigsaw + Google AI Team

Special Thanks to:
NIST + DARPA



Ensemble Model
Jigsaw

MODEL PREDICTION

Very likely manipulated

91% chance of manipulation

CONTRIBUTING DETECTORS:

Self-Consistency Splice
UC Berkeley

Splicebuster
University Federico II of Naples

Dense-Field Copy-Move
University Federico II of Naples

Color & Noise Patterns
University of Maryland

JPEG Dimples
UC Berkeley

Best detects: **Reformatting** in images.

Learn more

StyleGAN Detector
Jigsaw

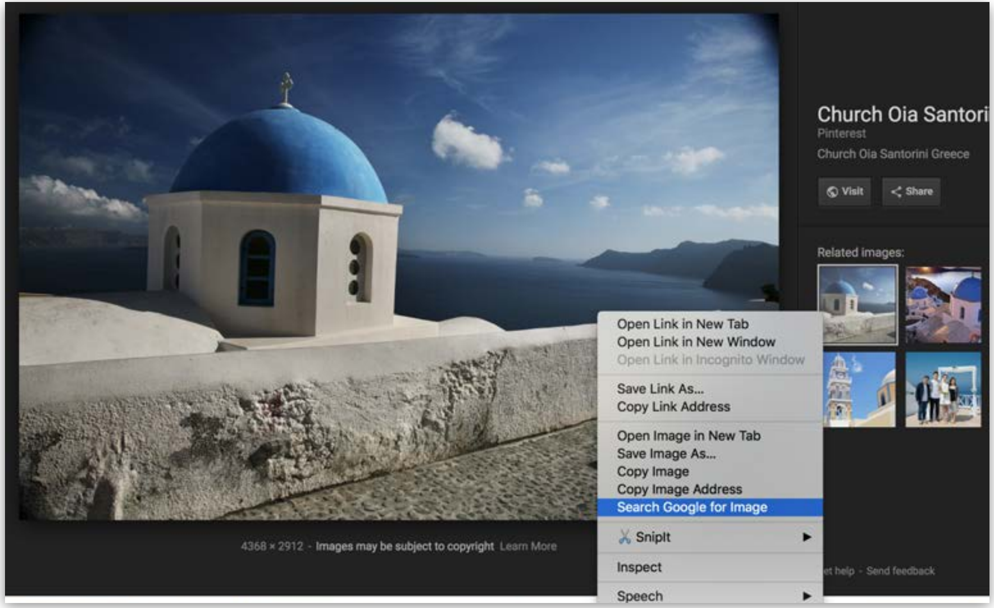
⊕ Add your detector

ASSEMBLER ALPHA **V 0.2**
Last updated: 25th Nov 2019

Google’s Media Integrity Efforts

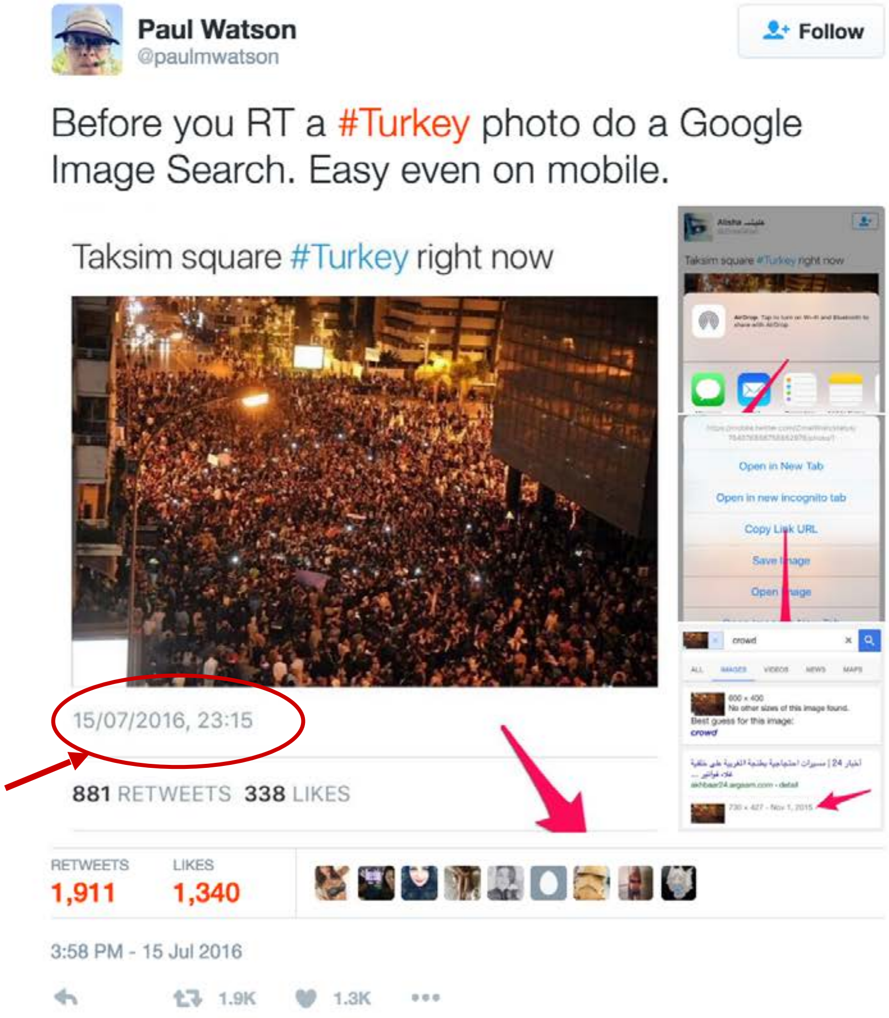
- Google News Initiative
- Fact Checking via ClaimReview/MediaReview
- Deepfake (audio + video) datasets
- CheapFakes/DeepFakes -> FactCheck/Journalists
- Provenance: GRIS
- Misleading / False Context
- Other research efforts

How to use Google Reverse Image Search (GRIS)?



1. Chrome: right-click on the image
2. Select “Search Google for Image”

Courtesy: Howard Zhou



Paul Watson
@paulmwatson

Before you RT a #Turkey photo do a Google Image Search. Easy even on mobile.

Taksim square #Turkey right now

15/07/2016, 23:15

881 RETWEETS 338 LIKES

Aisha عليشـ

Taksim square #Turkey right now

AirDrop. Tap to turn on Wi-Fi and Bluetooth to share with AirDrop.

600 × 400
No other sizes of this image found.
Best guess for this image:
crowd

أخبار 24 | مسيرات احتجاجية بطنجة المغربية على خلفية غلاء فواتير ...
akhbaar24.argaam.com › detail

730 × 427 - Nov 1, 2015

RETWEETS 1,911 LIKES 1,340

3:58 PM - 15 Jul 2016

#Turkey - July 15, 2016

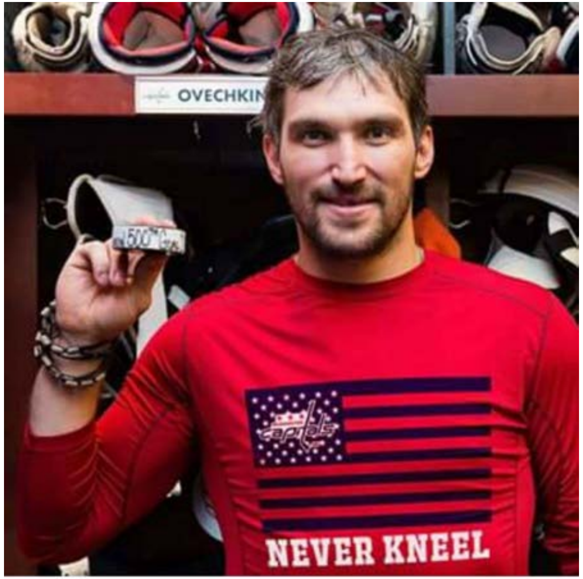
#Moroco - Nov 1, 2015

Google

Courtesy: Howard Zhou

VERIFY: The truth behind Ovechkin in the 'Never Kneel' shirt pic

“QUESTION: Did Alex Ovechkin wear a licensed Capitals shirt that says "Never Kneel?"
ANSWER: No, this photo is not legit.
SOURCES: Google Reverse Image Search”

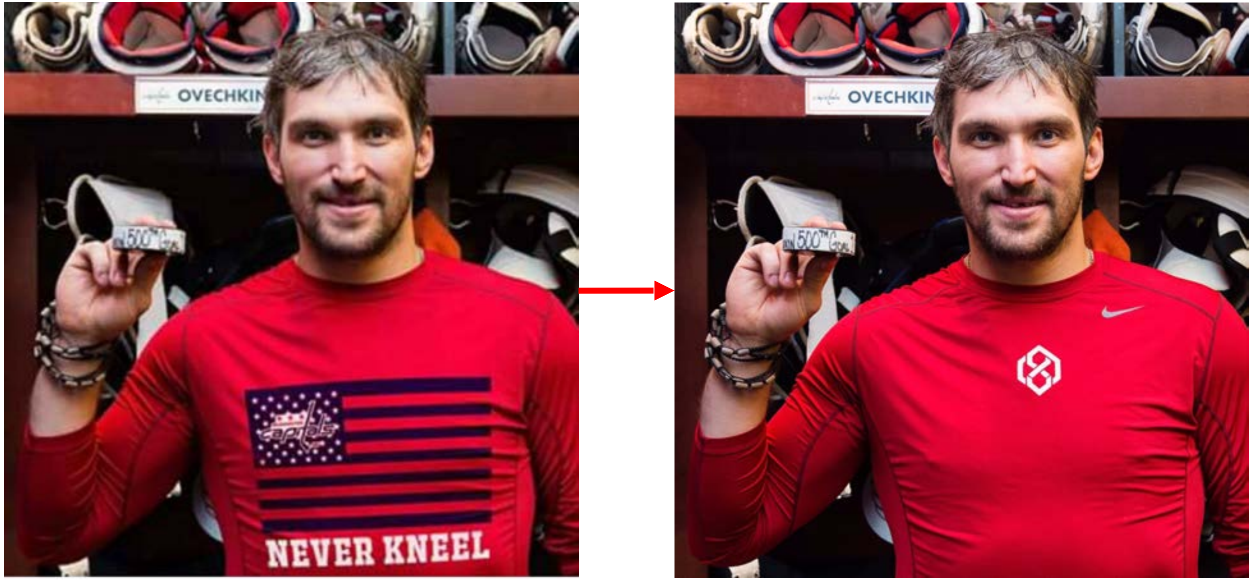


Google Source: wusa9.com

Courtesy: Howard Zhou

VERIFY: The truth behind Ovechkin in the 'Never Kneel' shirt pic

“QUESTION: Did Alex Ovechkin wear a licensed Capitals shirt that says "Never Kneel?"
ANSWER: No, this photo is not legit.
SOURCES: Google Reverse Image Search”



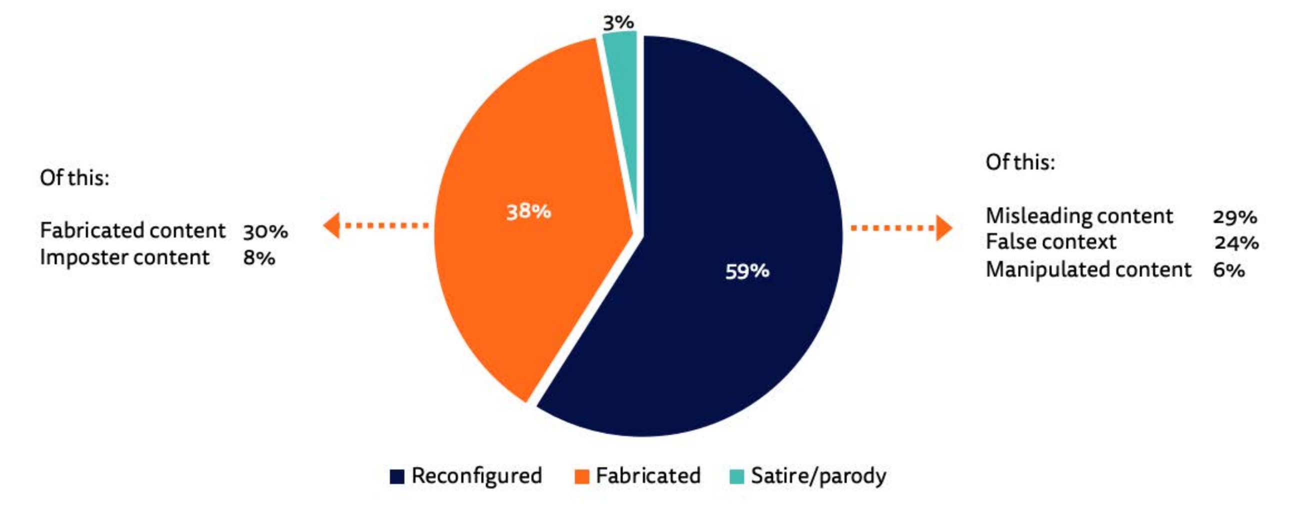
Google Source: wusa9.com

Courtesy: Howard Zhou

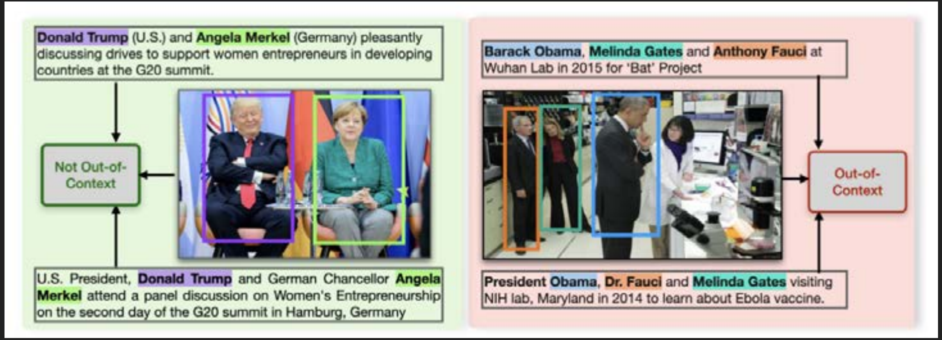
Oxford Reuters Institute

Types, Sources, and Claims of COVID-19 Misinformation

Authors: J. Scott Brennen, Felix M. Simon, Philip N. Howard, and Rasmus Kleis Nielsen



Out of context detection ?

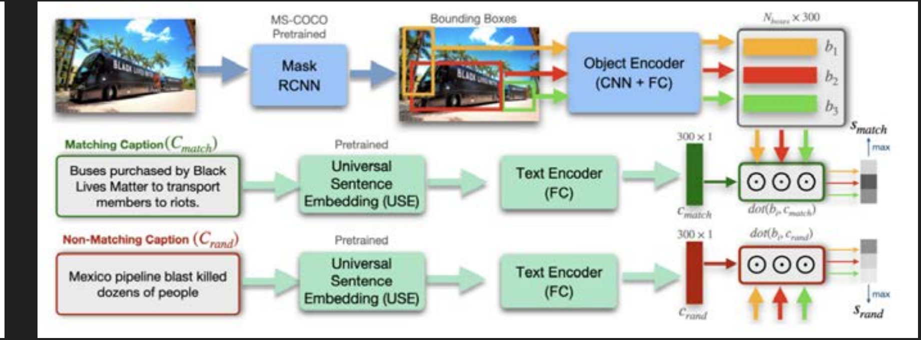


Donald Trump (U.S.) and Angela Merkel (Germany) pleasantly discussing drives to support women entrepreneurs in developing countries at the G20 summit.

U.S. President, Donald Trump and German Chancellor Angela Merkel attend a panel discussion on Women's Entrepreneurship on the second day of the G20 summit in Hamburg, Germany

Barack Obama, Melinda Gates and Anthony Fauci at Wuhan Lab in 2015 for 'Bat' Project

President Obama, Dr. Fauci and Melinda Gates visiting NIH lab, Maryland in 2014 to learn about Ebola vaccine.



Buses purchased by Black Lives Matter to transport members to riots.

Mexico pipeline blast killed dozens of people

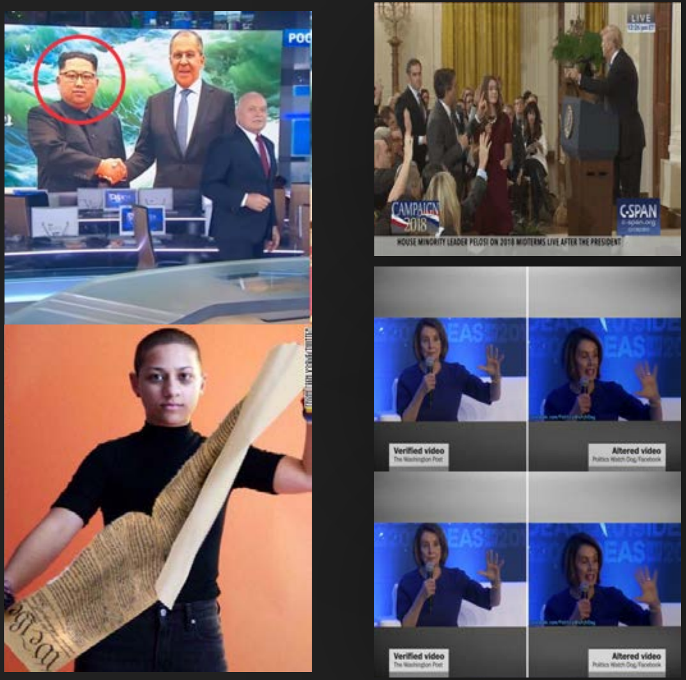
Detecting out-of-context image captions: with Shivangi Aneja, Matthias Niessner (TUM)

https://shivangi-aneja.github.io/projects/cosmos/

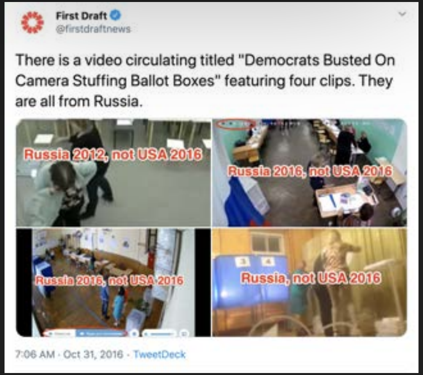
Grand Challenge on

Detecting Cheapfakes

Organized and sponsored by



Paid fake protesters were bussed in to the anti-Trump protests in Austin, Texas.



First Draft
@firstdraftnews

There is a video circulating titled "Democrats Busted On Camera Stuffing Ballot Boxes" featuring four clips. They are all from Russia.

Russia 2012, not USA 2016
Russia 2016, not USA 2016
Russia 2016, not USA 2016
Russia, not USA 2016

7:06 AM · Oct 31, 2016 · TweetDeck

DeepFakes, Liar’s Dividend

CheapFakes: Doctoring, Editing

CheapFakes: Change Context

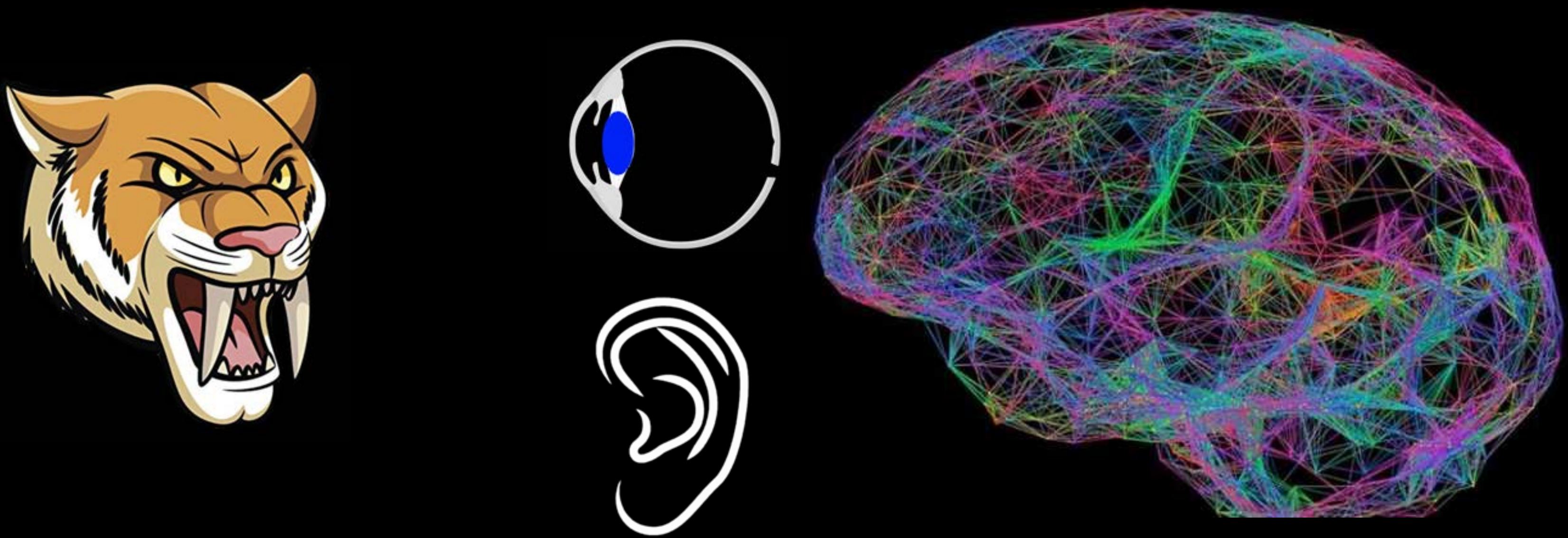
Fake Media ≠ Fake News

Most manipulations with DeepFakes or PhotoShop is innocent, for entertainment, or parody.

Malicious intent or real-world harm
-> Disinformation

Deciding what is mis/disinformation is a human task

World → Sensor → Truth ?



50,000 years of *Post-Truth*

Story Telling

Memory
& Beliefs

No Silver Bullet: Multidisciplinary Effort