Everything in your archive is now fake.

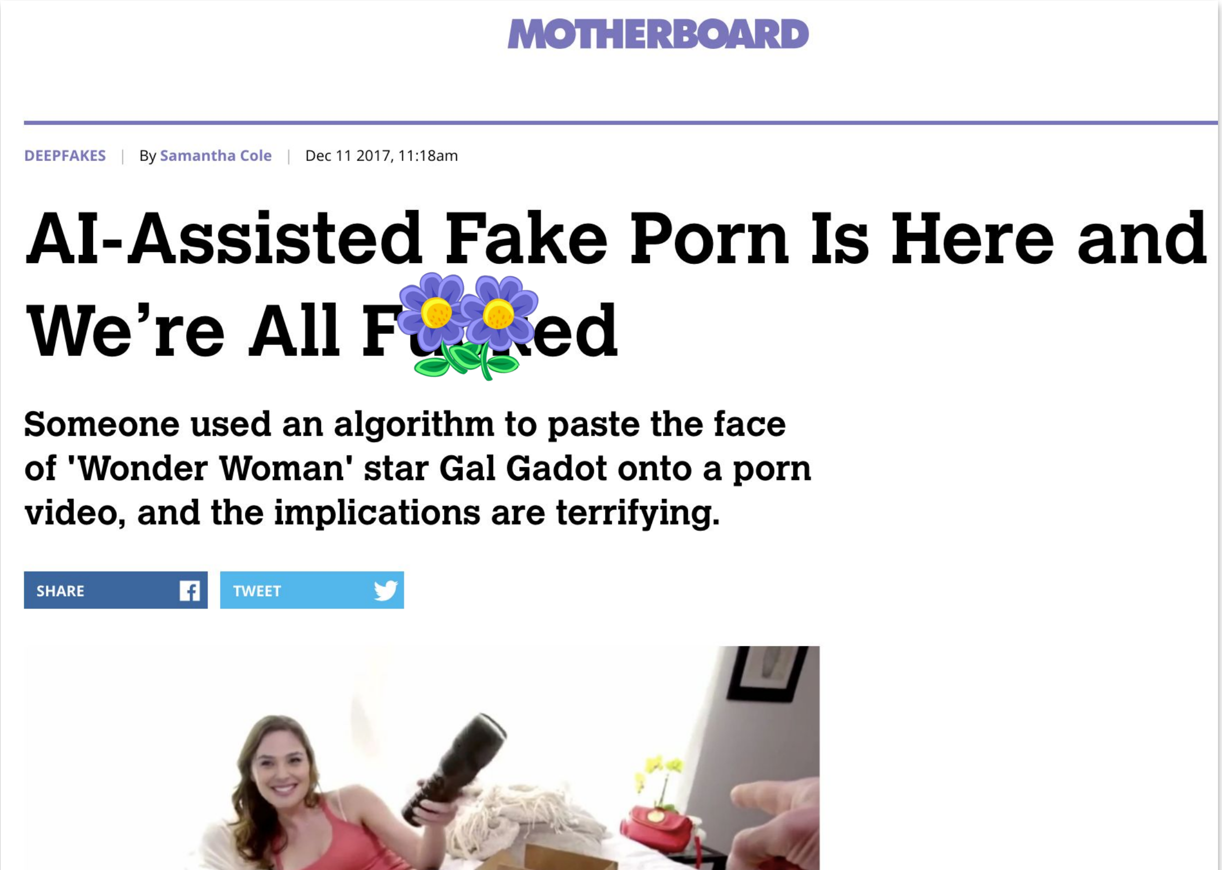
Gaurav Oberoi
Allen Institute for AI
#AMIAConference2018

Back on Dec 11, 2017

Back on Dec 11, 2017

Back on Dec 11, 2017

Suddenly, this happened:



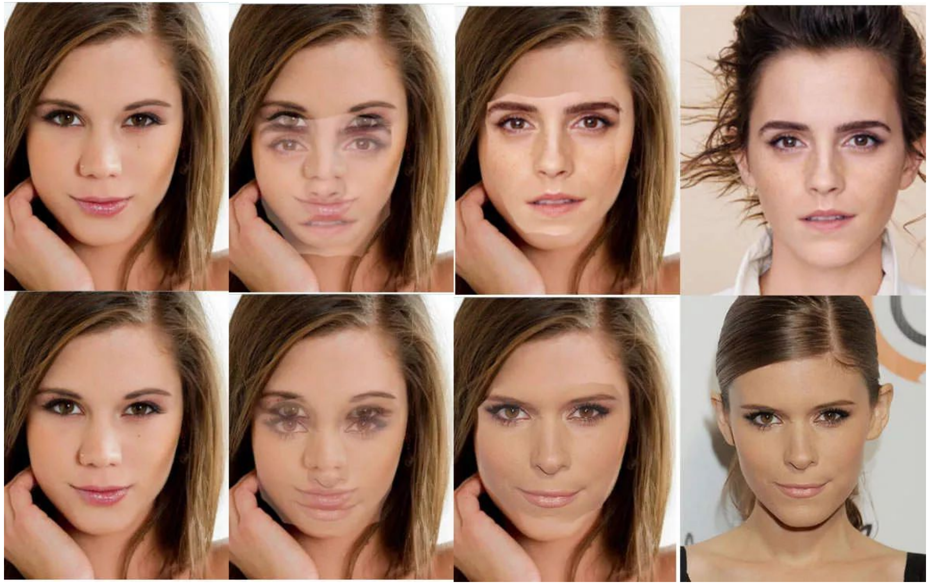
MOTHERBOARD

DEEPFAKES | By Samantha Cole | Dec 11 2017, 11:18am

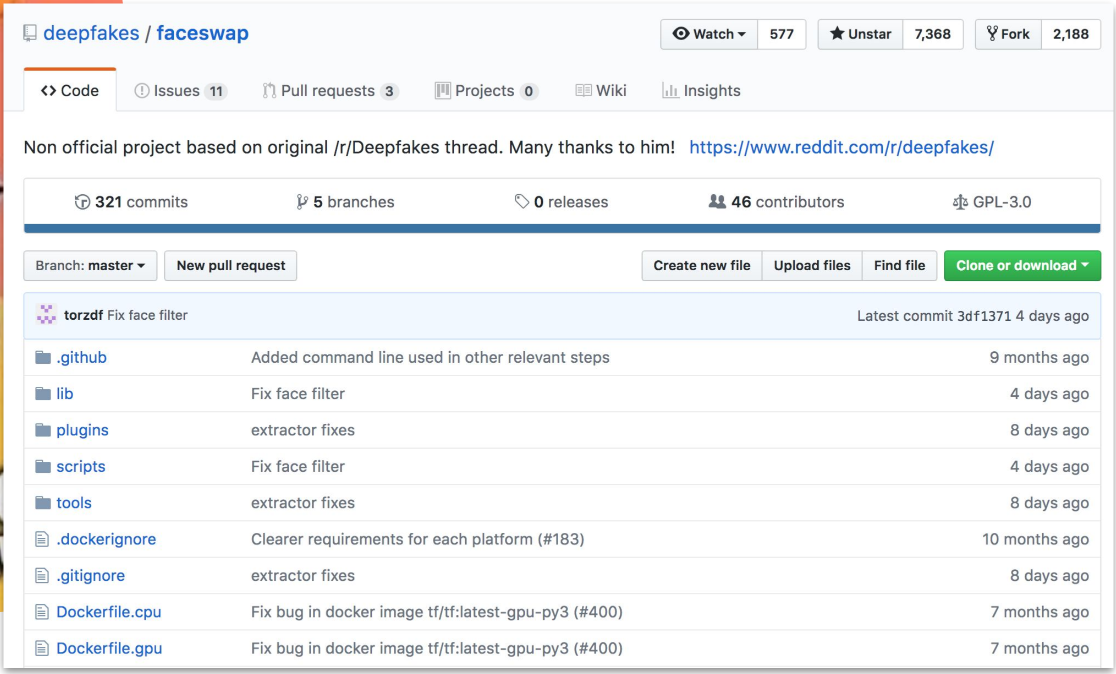
AI-Assisted Fake Porn Is Here and We’re All F🌼🌼ed

Someone used an algorithm to paste the face of 'Wonder Woman' star Gal Gadot onto a porn video, and the implications are terrifying.

A Reddit user "Deepfakes" started posting AI driven fake celebrity NSFW videos



They also released the code. You needed to be a coder to use it.


deepfakes / faceswap

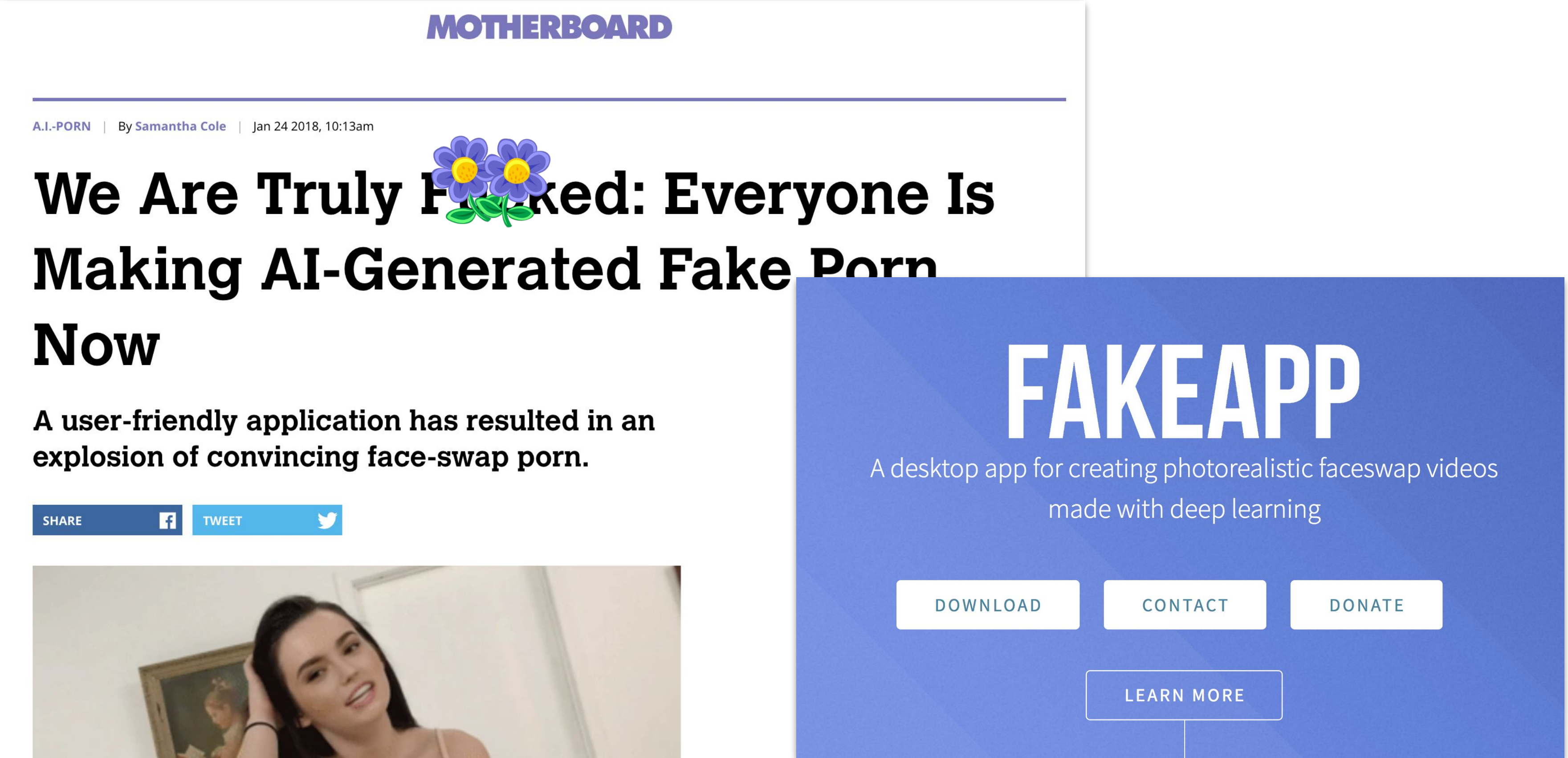
Non official project based on original /r/Deepfakes thread. Many thanks to him! https://www.reddit.com/r/deepfakes/

321 commits | 5 branches | 0 releases | 46 contributors | GPL-3.0

torzdf Fix face filter — Latest commit 3df1371 4 days ago

.github	Added command line used in other relevant steps	9 months ago
lib	Fix face filter	4 days ago
plugins	extractor fixes	8 days ago
scripts	Fix face filter	4 days ago
tools	extractor fixes	8 days ago
.dockerignore	Clearer requirements for each platform (#183)	10 months ago
.gitignore	extractor fixes	8 days ago
Dockerfile.cpu	Fix bug in docker image tf/tf:latest-gpu-py3 (#400)	7 months ago
Dockerfile.gpu	Fix bug in docker image tf/tf:latest-gpu-py3 (#400)	7 months ago

But then, an app came out, so anyone could do it



MOTHERBOARD

A.I.-PORN | By Samantha Cole | Jan 24 2018, 10:13am

We Are Truly F🌼🌼ked: Everyone Is Making AI-Generated Fake Porn Now

A user-friendly application has resulted in an explosion of convincing face-swap porn.

SHARE TWEET

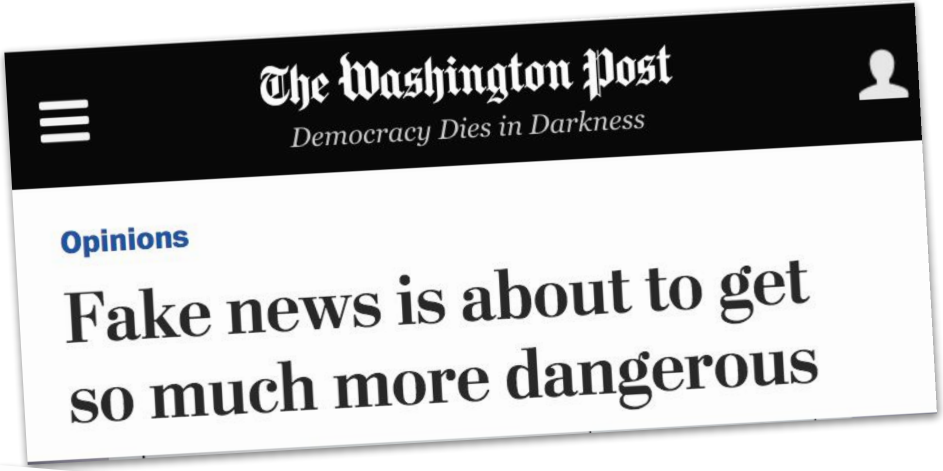
FAKEAPP

A desktop app for creating photorealistic faceswap videos made with deep learning

DOWNLOAD CONTACT DONATE

LEARN MORE

And everybody flipped out

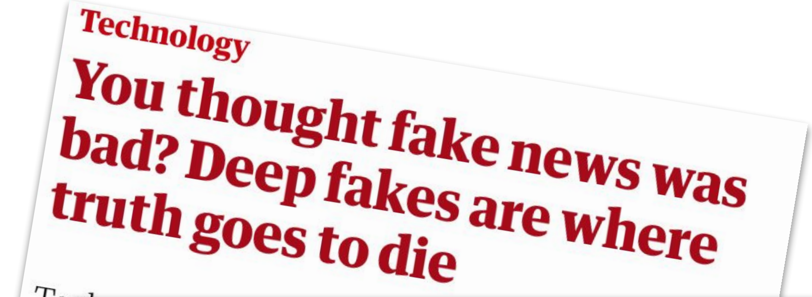


The Washington Post

Democracy Dies in Darkness

Opinions

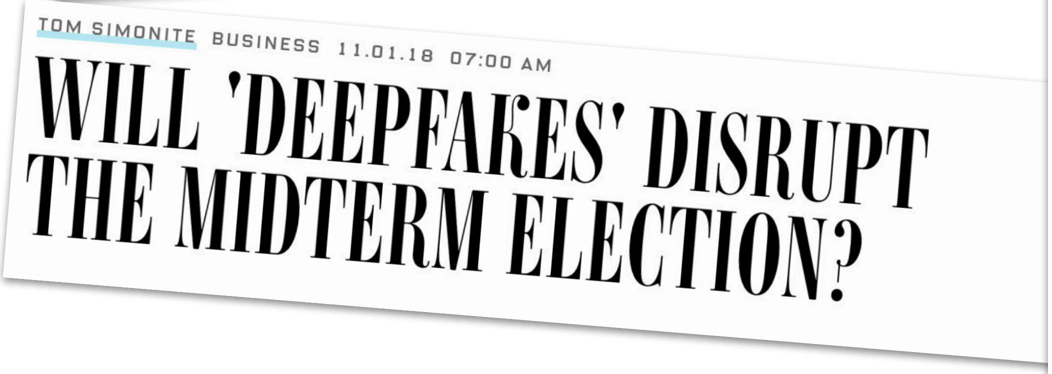
Fake news is about to get so much more dangerous



Technology

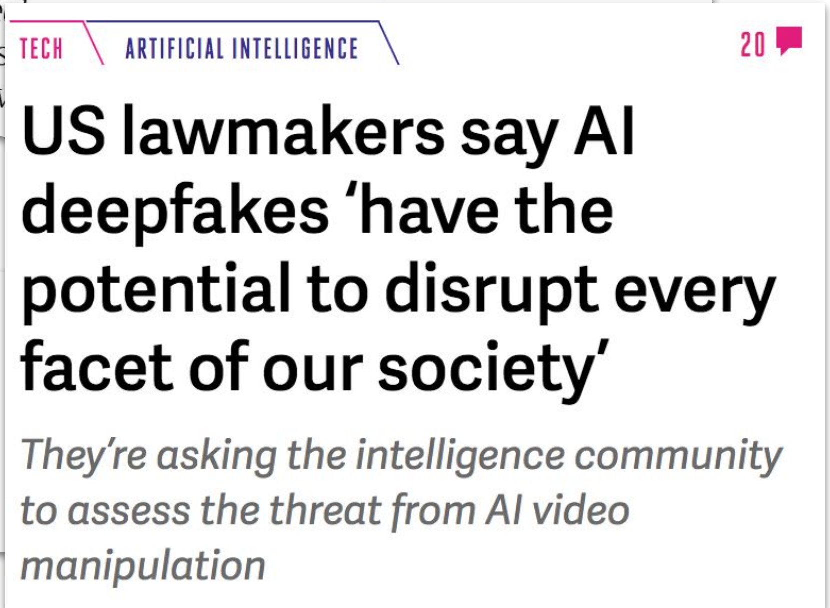
You thought fake news was bad? Deep fakes are where truth goes to die

Te
has
wav



TOM SIMONITE BUSINESS 11.01.18 07:00 AM

WILL 'DEEPFAKES' DISRUPT THE MIDTERM ELECTION?

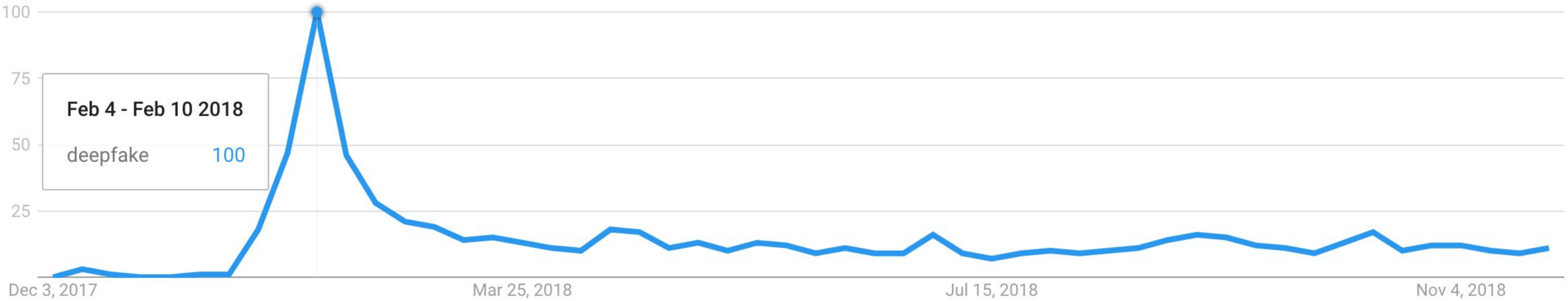


TECH ARTIFICIAL INTELLIGENCE 20

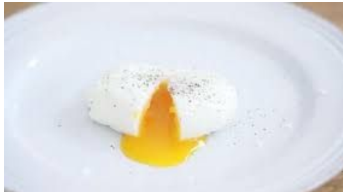
US lawmakers say AI deepfakes 'have the potential to disrupt every facet of our society'

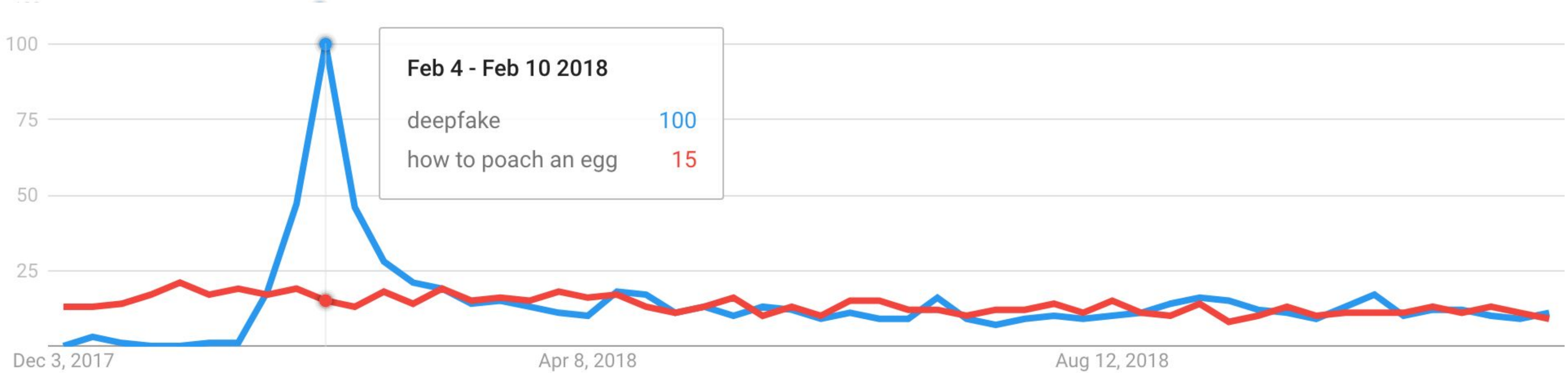
They're asking the intelligence community to assess the threat from AI video manipulation

Feb 2018 is when searches peaked for: "deepfake" (blue line)



That’s about as popular as “how to poach an egg” (red line)

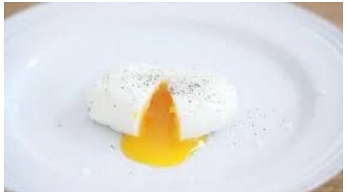


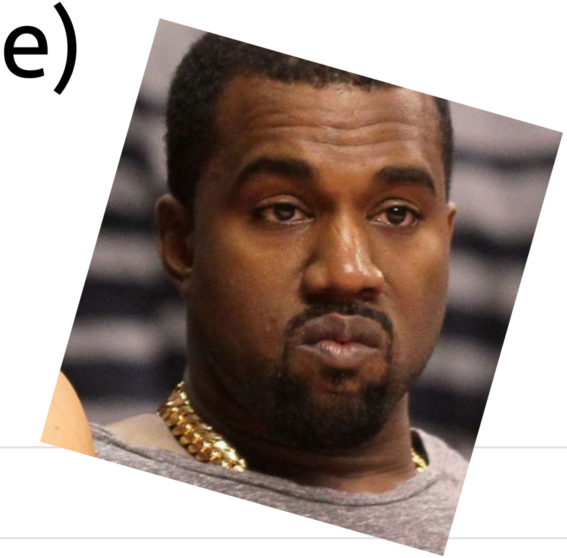


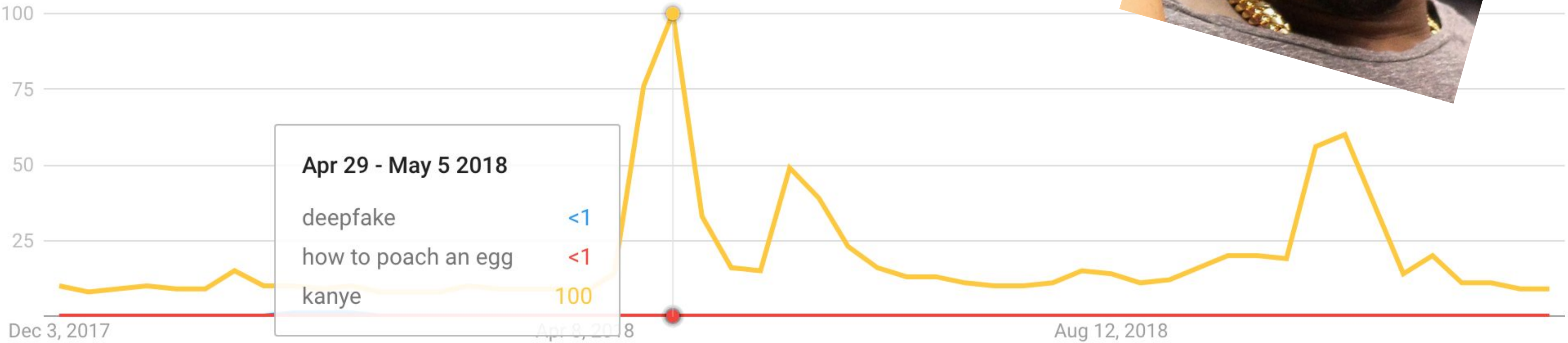
Feb 4 - Feb 10 2018

deepfake	100
how to poach an egg	15

And literally 20-100X less popular than: “Kanye” (yellow line)



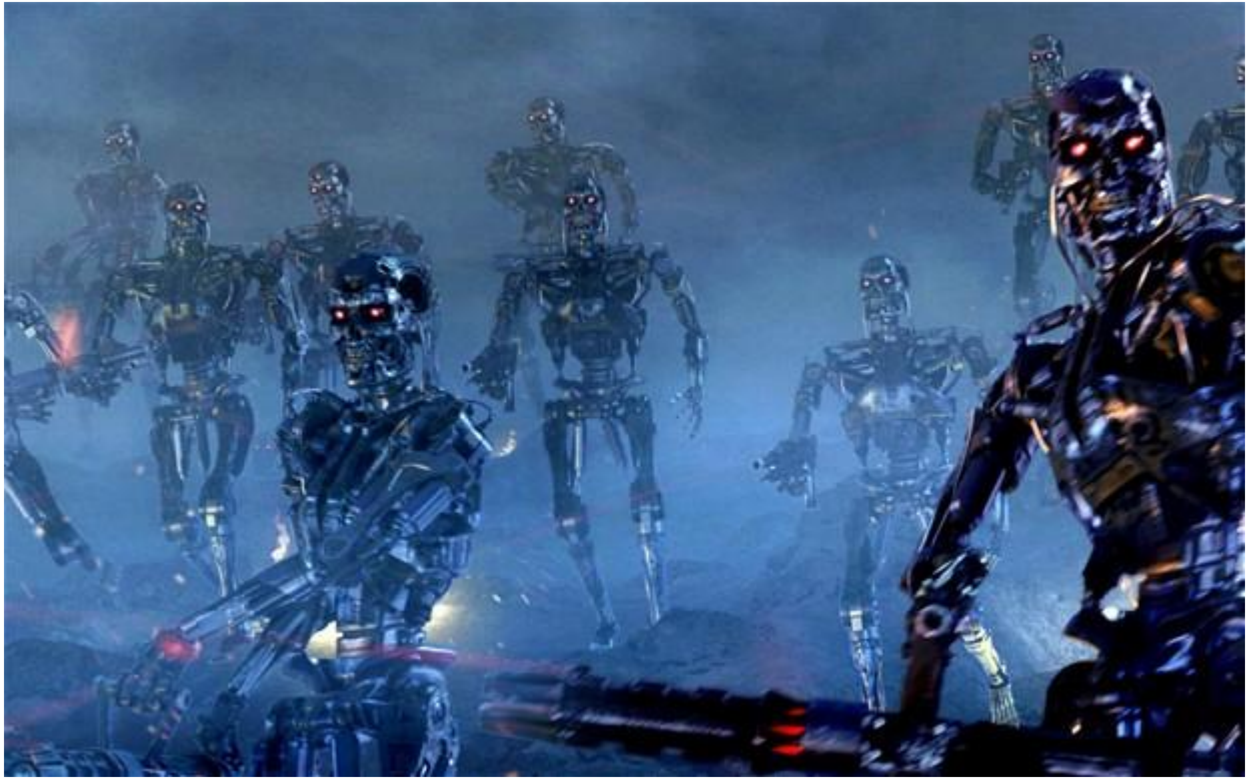




Apr 29 - May 5 2018

deepfake	<1
how to poach an egg	<1
kanye	100

Hyper realistic AI generated synthetic media is coming at an accelerating pace



And it will start to impact your day job, much like fake news

My goals in this session:

- ☑ Raise awareness of AI generated synthetic media
- ☑ Suggest coping tools
- ☑ Make superfluous use of gifs

But first, why am I giving this talk?

I come from an engineering/startup background, not AI research

amazon

Engineering peon

PrecisionPolling

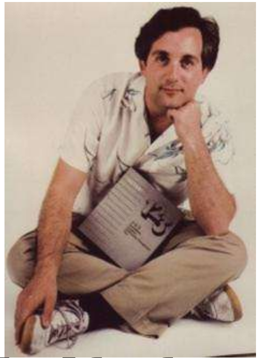
Started and sold another startup

BillMonk

Started and sold a startup

SurveyMonkey®

Helped grow from 50 to 700 people



Worked for founder of Lotus 123

Built one of Seattle's earliest startup communities

Currently, my day job is to commercialize AI research:

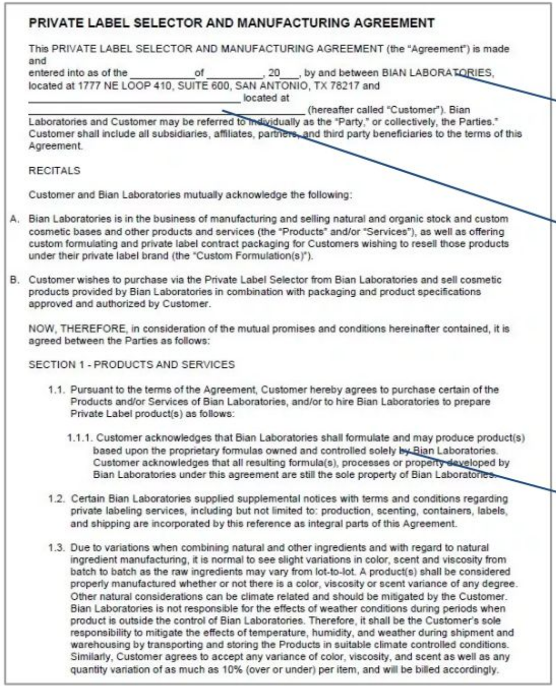
AI2

ALLEN INSTITUTE
for ARTIFICIAL INTELLIGENCE

In the last year, I've worked on several applications of AI:

Analyzing contracts to extract meaning

DocuSmart



PRIVATE LABEL SELECTOR AND MANUFACTURING AGREEMENT

This PRIVATE LABEL SELECTOR AND MANUFACTURING AGREEMENT (the "Agreement") is made and
entered into as of the ____________of __________, 20___, by and between BIAN LABORATORIES, located at 1777 NE LOOP 410, SUITE 600, SAN ANTONIO, TX 78217 and
__________________________________ located at
____ (hereafter called "Customer"). Bian Laboratories and Customer may be referred to individually as the "Party," or collectively, the Parties." Customer shall include all subsidiaries, affiliates, partners, and third party beneficiaries to the terms of this Agreement.

RECITALS

Customer and Bian Laboratories mutually acknowledge the following:

A. Bian Laboratories is in the business of manufacturing and selling natural and organic stock and custom cosmetic bases and other products and services (the "Products" and/or "Services"), as well as offering custom formulating and private label contract packaging for Customers wishing to resell those products under their private label brand (the "Custom Formulation(s)").

B. Customer wishes to purchase via the Private Label Selector from Bian Laboratories and sell cosmetic products provided by Bian Laboratories in combination with packaging and product specifications approved and authorized by Customer.

NOW, THEREFORE, in consideration of the mutual promises and conditions hereinafter contained, it is agreed between the Parties as follows:

SECTION 1 - PRODUCTS AND SERVICES

1.1. Pursuant to the terms of the Agreement, Customer hereby agrees to purchase certain of the Products and/or Services of Bian Laboratories, and/or to hire Bian Laboratories to prepare Private Label product(s) as follows:

1.1.1. Customer acknowledges that Bian Laboratories shall formulate and may produce product(s) based upon the proprietary formulas owned and controlled solely by Bian Laboratories. Customer acknowledges that all resulting formula(s), processes or property developed by Bian Laboratories under this agreement are still the sole property of Bian Laboratories.

1.2. Certain Bian Laboratories supplied supplemental notices with terms and conditions regarding private labeling services, including but not limited to: production, scenting, containers, labels, and shipping are incorporated by this reference as integral parts of this Agreement.

1.3. Due to variations when combining natural and other ingredients and with regard to natural ingredient manufacturing, it is normal to see slight variations in color, scent and viscosity from batch to batch as the raw ingredients may vary from lot-to-lot. A product(s) shall be considered properly manufactured whether or not there is a color, viscosity or scent variance of any degree. Other natural considerations can be climate related and should be mitigated by the Customer. Bian Laboratories is not responsible for the effects of weather conditions during periods when product is outside the control of Bian Laboratories. Therefore, it shall be the Customer's sole responsibility to mitigate the effects of temperature, humidity, and weather during shipment and warehousing by transporting and storing the Products in suitable climate controlled conditions. Similarly, Customer agrees to accept any variance of color, viscosity, and scent as well as any quantity variation of as much as 10% (over or under) per item, and will be billed accordingly.

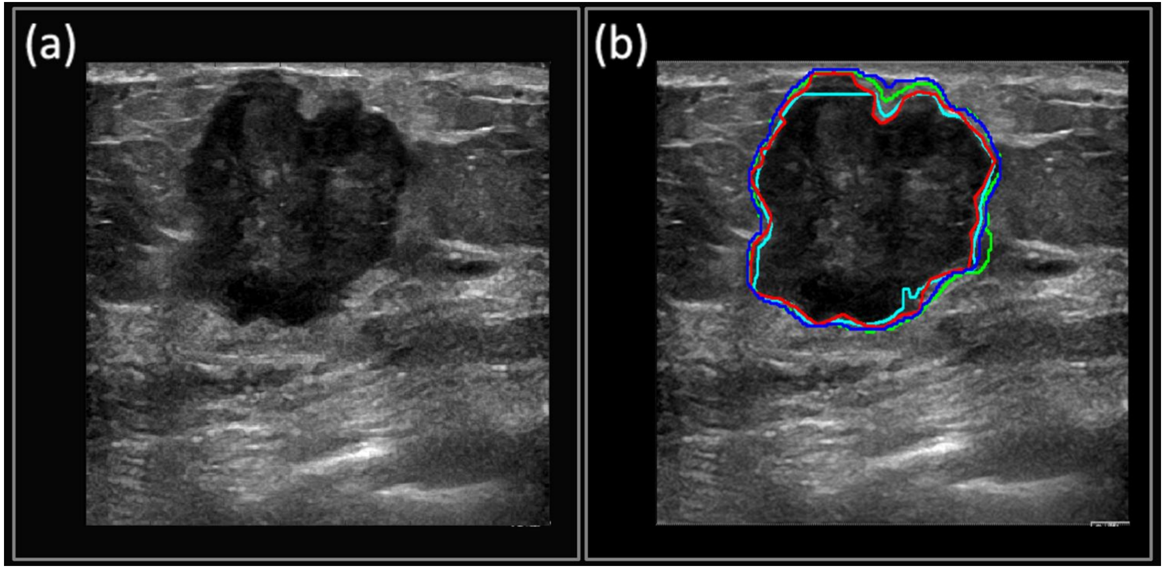
Vendor:
Bain Laboratories

Term Dates:
2018-2020

Items:
Enumeration of products and prices.

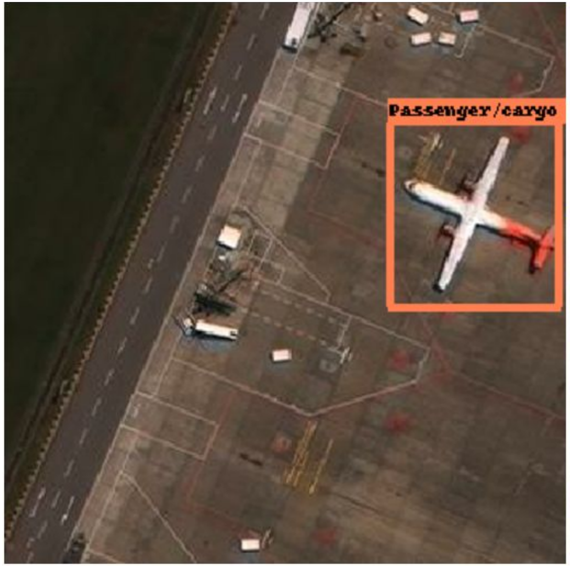
In the last year, I’ve worked on several applications of AI:

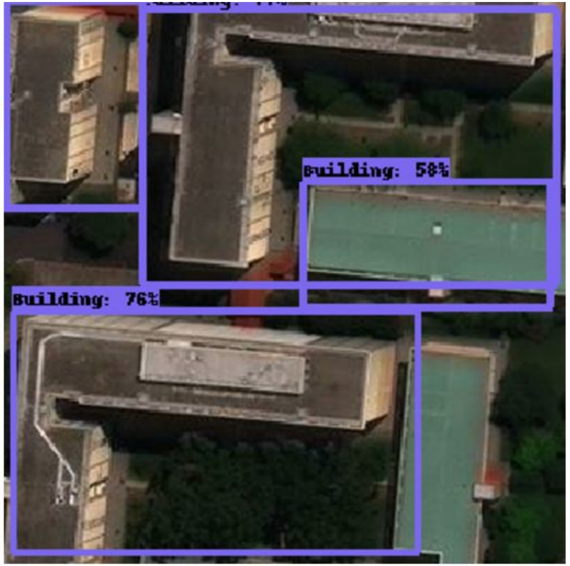
Finding anatomical parts in ultrasound images



In the last year, I’ve worked on several applications of AI:

Identifying vehicles and buildings in satellite imagery

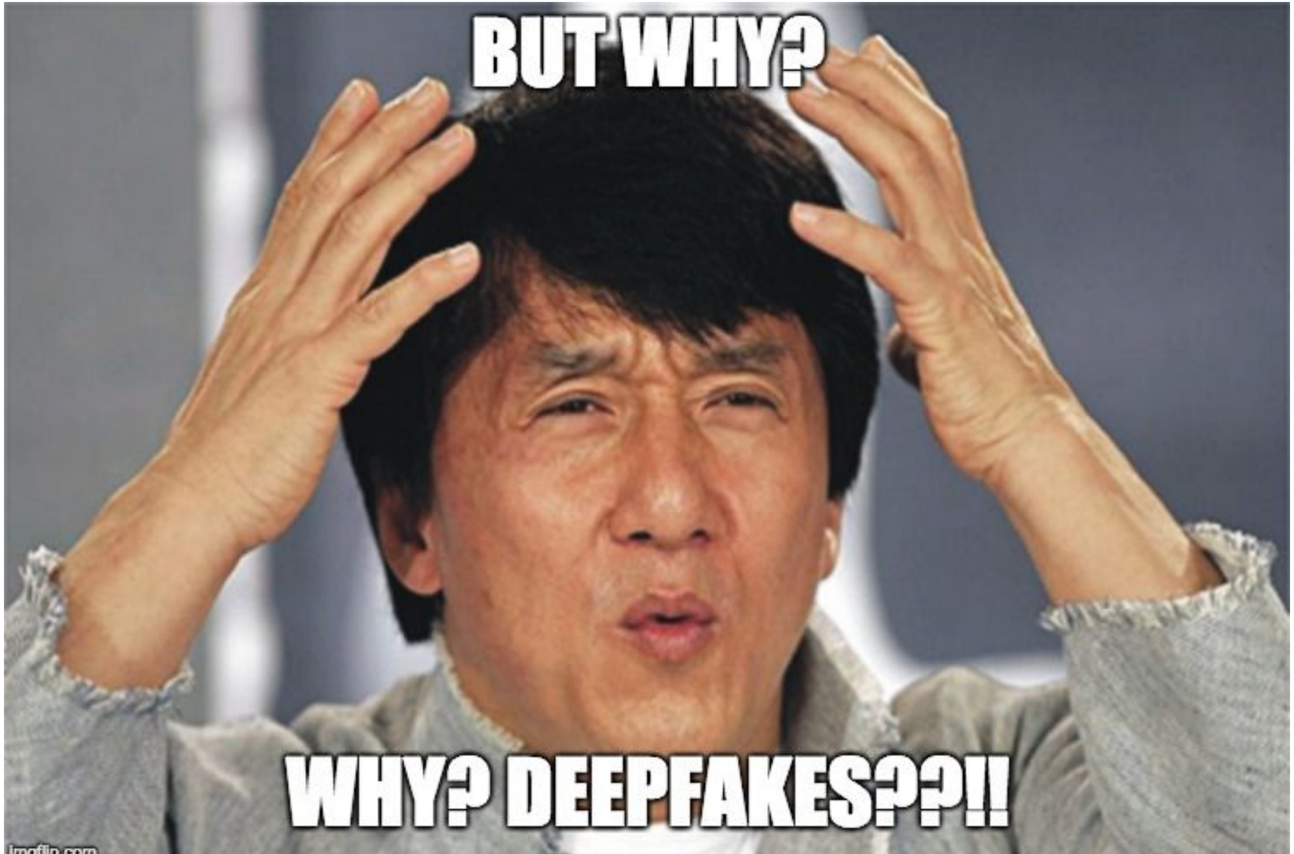




In the last year, I’ve worked on several applications of AI:

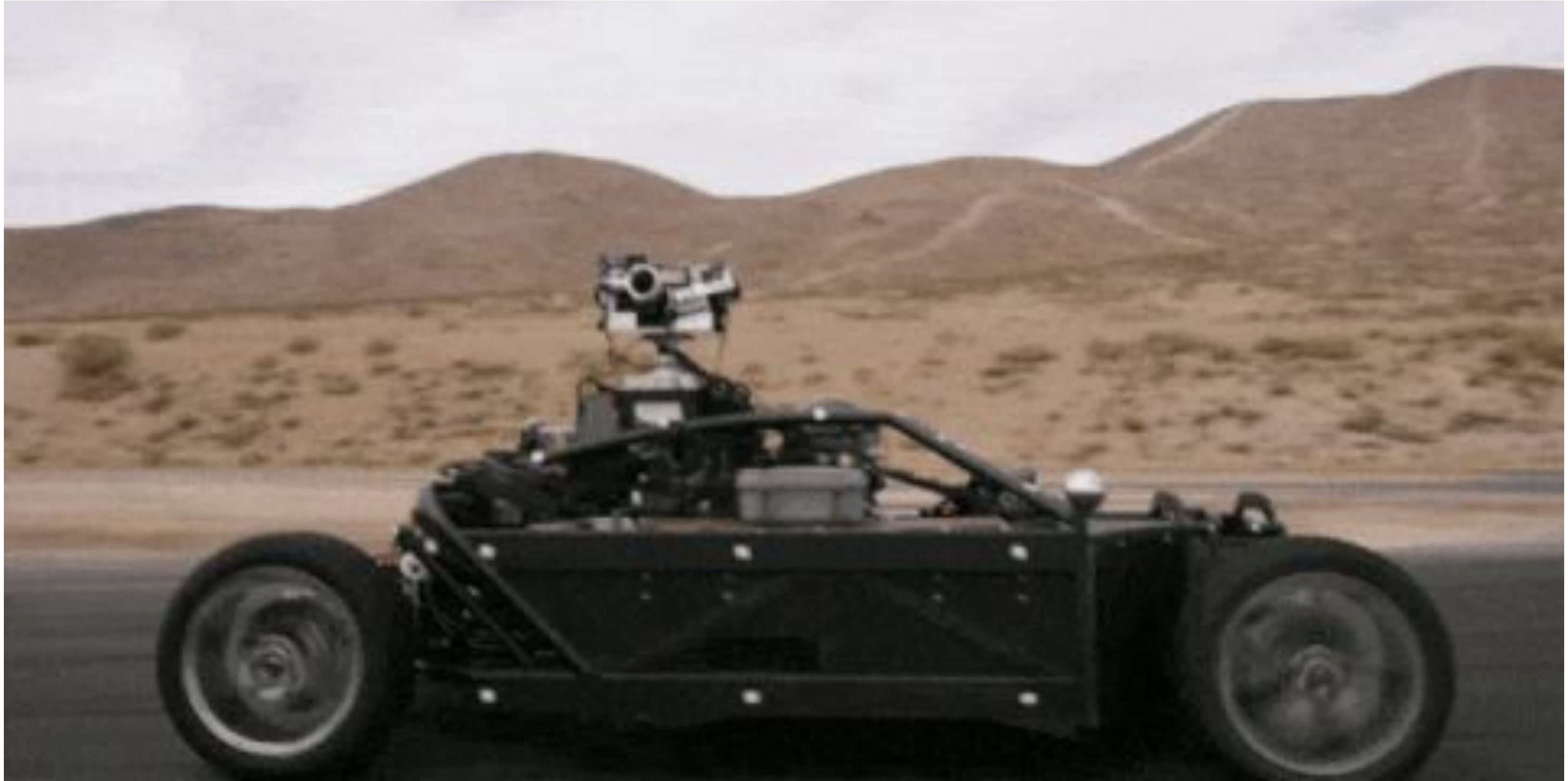
And AI generated synthetic media

I know what you're thinking…

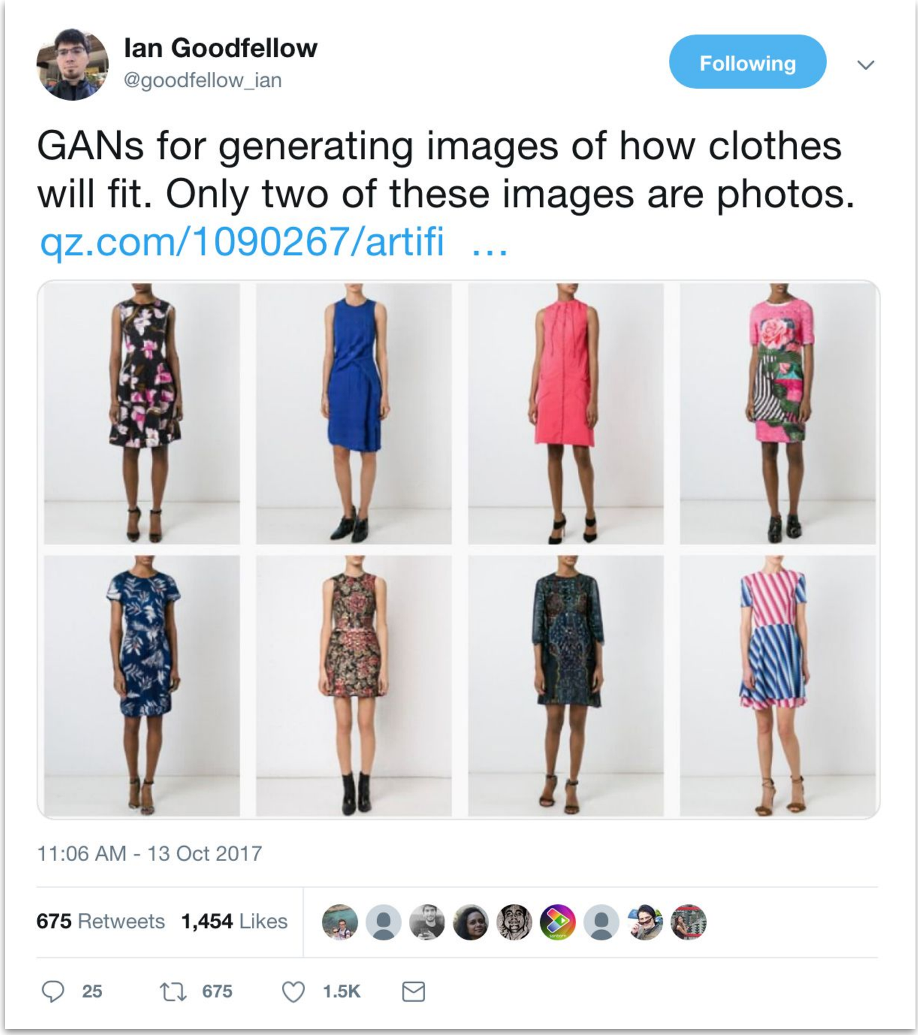


Large computer graphics market, e.g.: 75% of Ikea’s catalog is rendered (2014)

Most car commercials are shot without the car (2016)



Retail ready images produced from photos of cloth (2018)



Ian Goodfellow
@goodfellow_ian

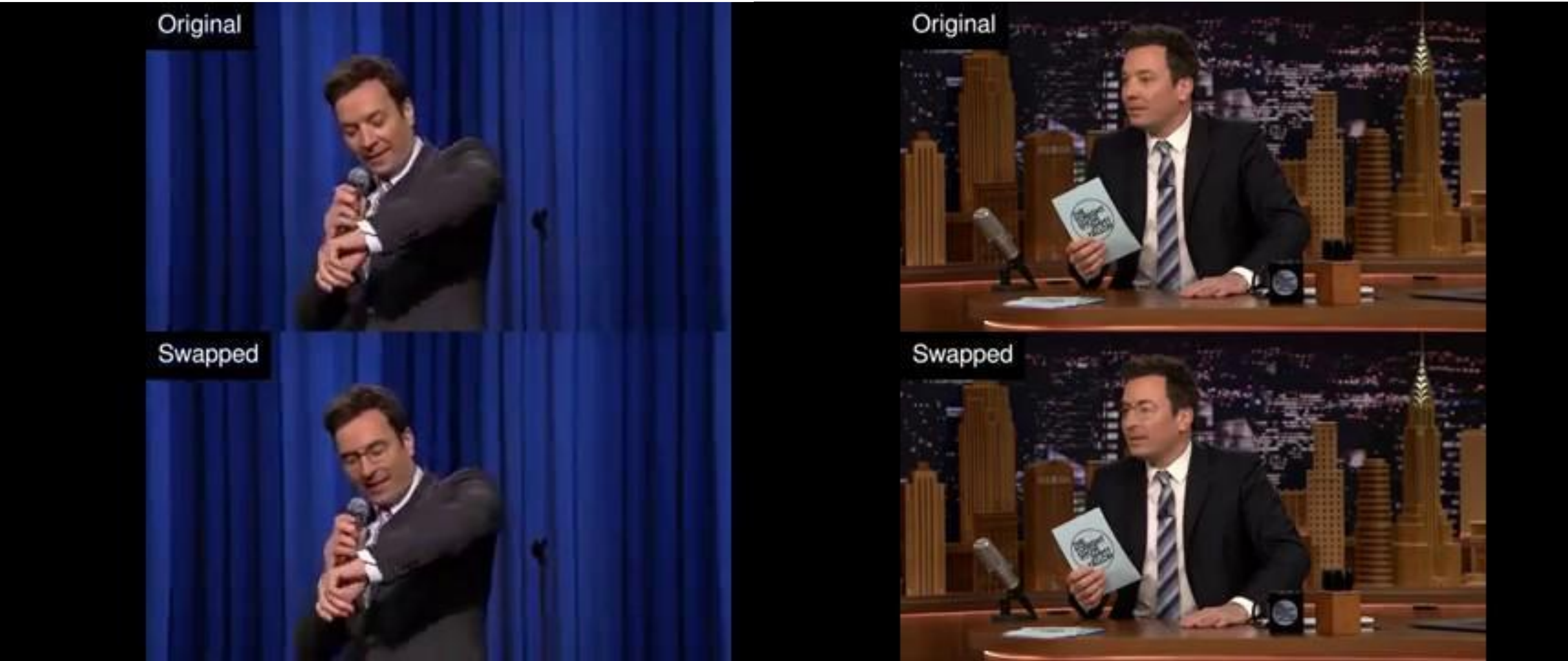
GANs for generating images of how clothes will fit. Only two of these images are photos. qz.com/1090267/artifi …

11:06 AM - 13 Oct 2017

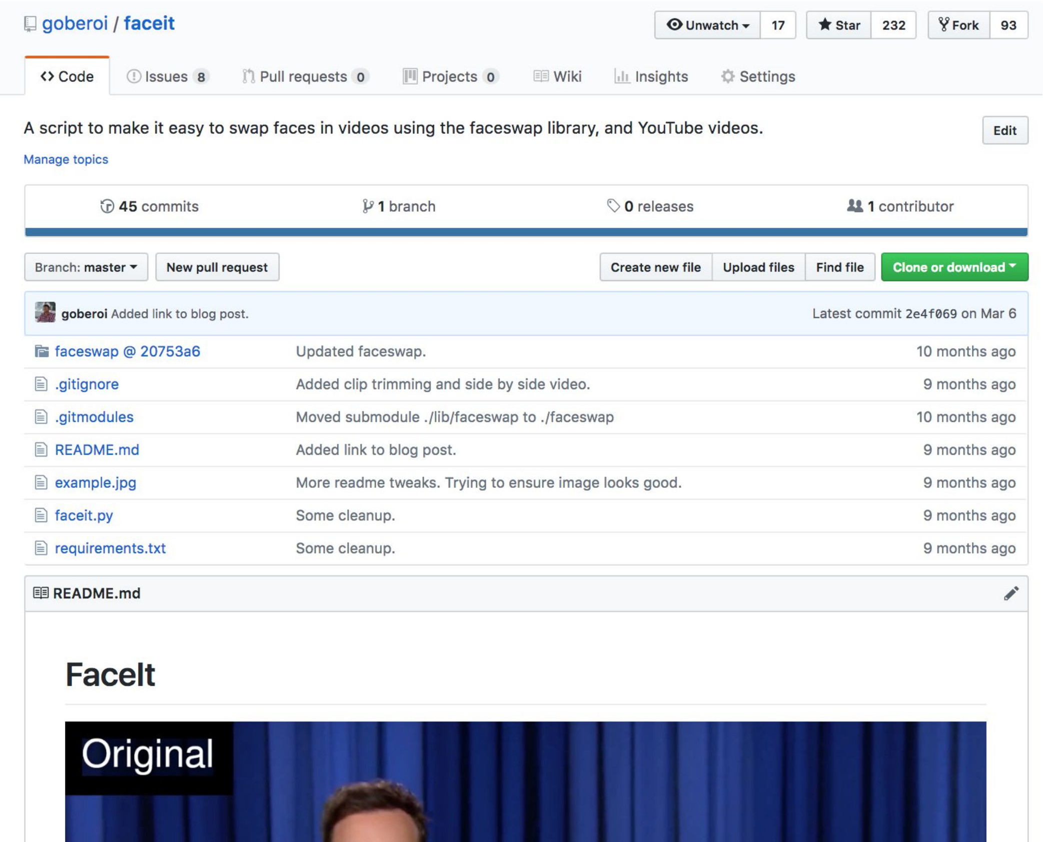
675 Retweets 1,454 Likes

Ok… so why am I giving this talk?

Created some bumpin' demos swapping the faces of two TV hosts



Released the tooling I built on top of existing Deepfakes code



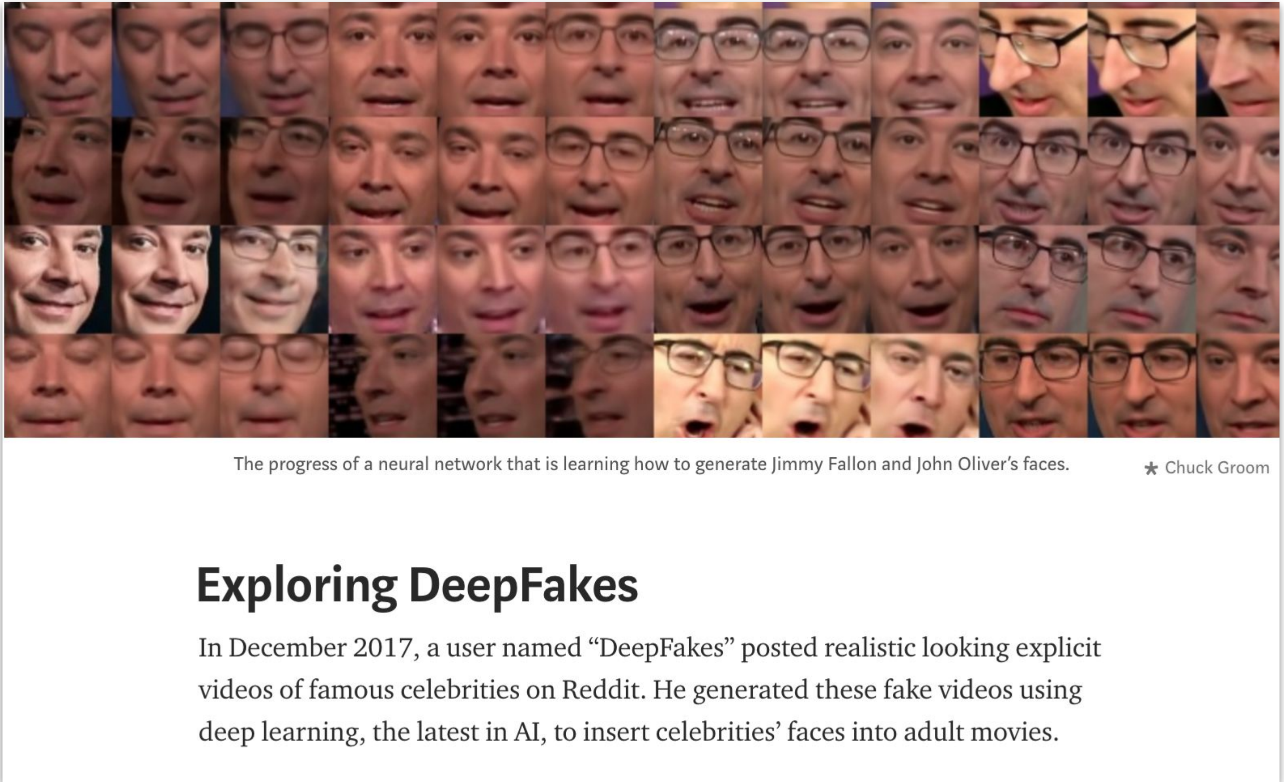
A script to make it easy to swap faces in videos using the faceswap library, and YouTube videos.

goberoi Added link to blog post. Latest commit 2e4f069 on Mar 6

faceswap @ 20753a6	Updated faceswap.	10 months ago
.gitignore	Added clip trimming and side by side video.	9 months ago
.gitmodules	Moved submodule ./lib/faceswap to ./faceswap	10 months ago
README.md	Added link to blog post.	9 months ago
example.jpg	More readme tweaks. Trying to ensure image looks good.	9 months ago
faceit.py	Some cleanup.	9 months ago
requirements.txt	Some cleanup.	9 months ago

FaceIt

And wrote an informative post explaining it all, and more



The progress of a neural network that is learning how to generate Jimmy Fallon and John Oliver's faces.

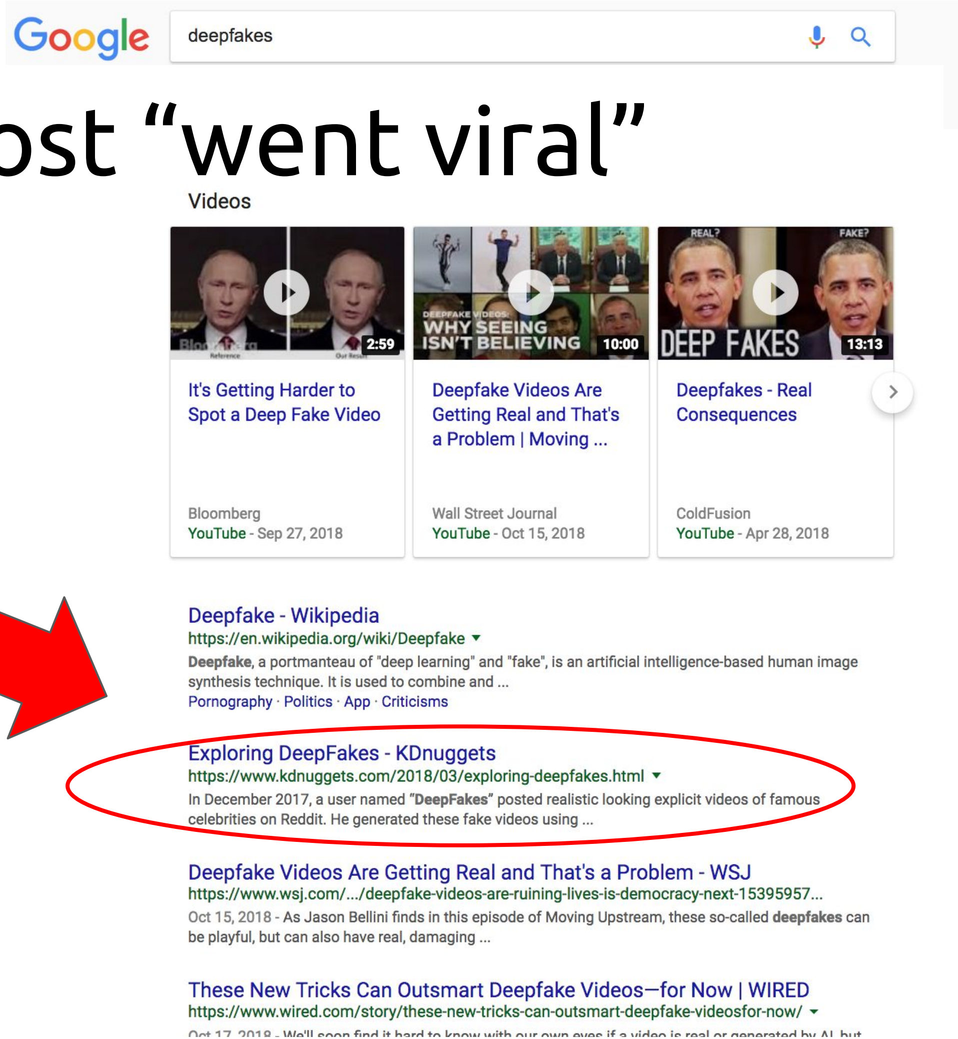
★ Chuck Groom

Exploring DeepFakes

In December 2017, a user named "DeepFakes" posted realistic looking explicit videos of famous celebrities on Reddit. He generated these fake videos using deep learning, the latest in AI, to insert celebrities' faces into adult movies.

And suddenly, my post “went viral”

#2 on Google for “Deepfakes”



deepfakes

Videos

It's Getting Harder to Spot a Deep Fake Video
Bloomberg
YouTube - Sep 27, 2018

Deepfake Videos Are Getting Real and That's a Problem | Moving ...
Wall Street Journal
YouTube - Oct 15, 2018

Deepfakes - Real Consequences
ColdFusion
YouTube - Apr 28, 2018

Deepfake - Wikipedia
https://en.wikipedia.org/wiki/Deepfake
Deepfake, a portmanteau of "deep learning" and "fake", is an artificial intelligence-based human image synthesis technique. It is used to combine and ...
Pornography · Politics · App · Criticisms

Exploring DeepFakes - KDnuggets
https://www.kdnuggets.com/2018/03/exploring-deepfakes.html
In December 2017, a user named **"DeepFakes"** posted realistic looking explicit videos of famous celebrities on Reddit. He generated these fake videos using ...

Deepfake Videos Are Getting Real and That's a Problem - WSJ
https://www.wsj.com/.../deepfake-videos-are-ruining-lives-is-democracy-next-15395957...
Oct 15, 2018 - As Jason Bellini finds in this episode of Moving Upstream, these so-called **deepfakes** can be playful, but can also have real, damaging ...

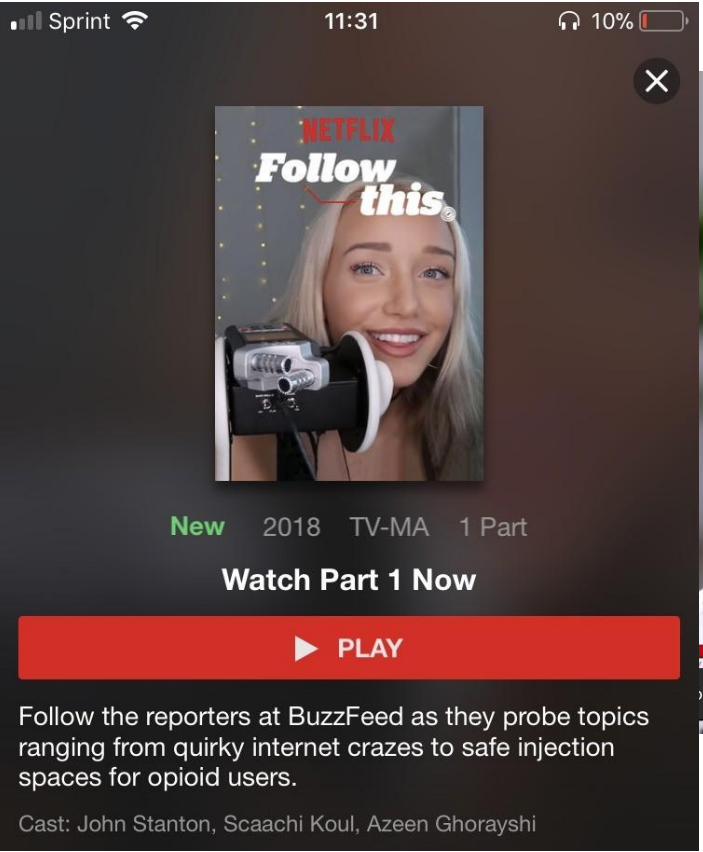
These New Tricks Can Outsmart Deepfake Videos—for Now | WIRED
https://www.wired.com/story/these-new-tricks-can-outsmart-deepfake-videosfor-now/
Oct 17, 2018 - We'll soon find it hard to know with our own eyes if a video is real or generated by AI, but

Journalists started to call…

Yes, it's a problem, but let's stay calm

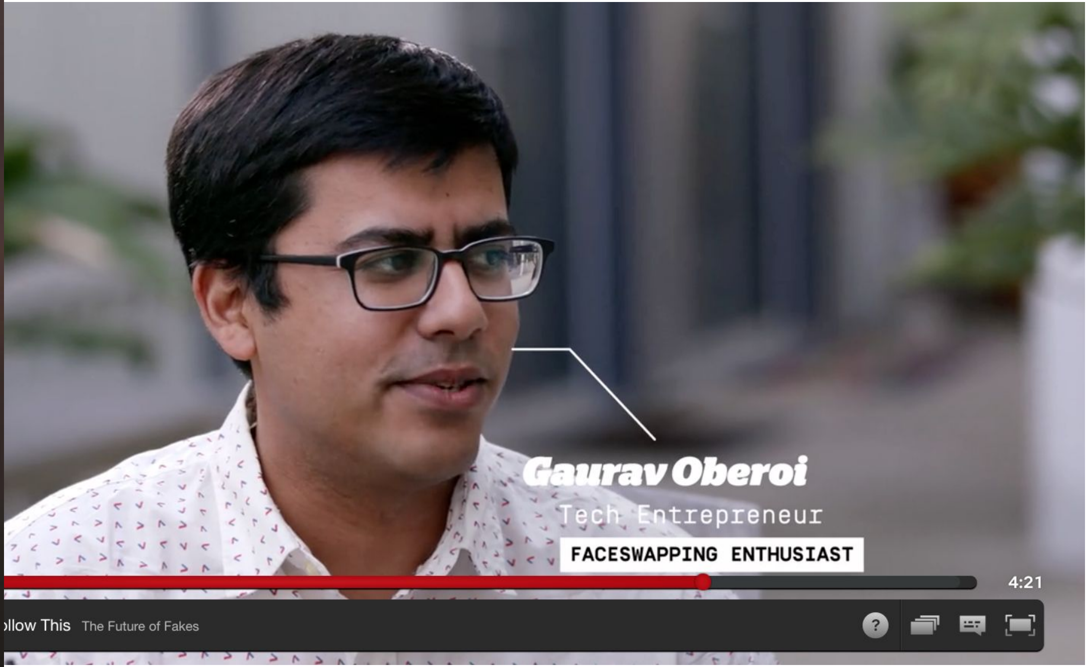
I see, so you're saying we're **doomed**?

And Buzzfeed asked me to comment in their Netflix documentary:



Follow the reporters at BuzzFeed as they probe topics ranging from quirky internet crazes to safe injection spaces for opioid users.

Cast: John Stanton, Scaachi Koul, Azeen Ghorayshi



And got me a speaking slot at AMIA Conf 2018!

❤ AMIA

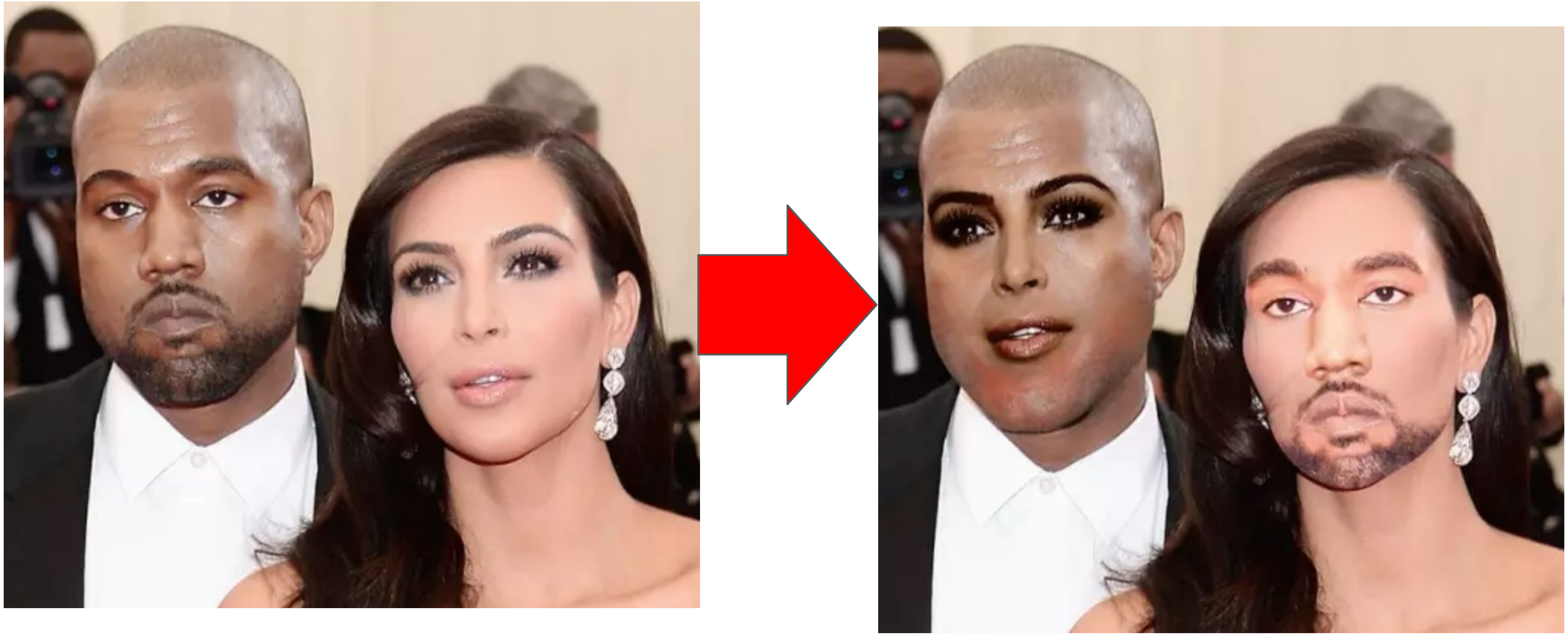
And 2 days ago, Fallon and Oliver saw it!

In, summary:

I am not an expert, but have some experience with this technology.

So what exactly are Deepfakes?

It’s not your gramma’s face swap



There are plenty of apps for that

There are plenty of apps for that

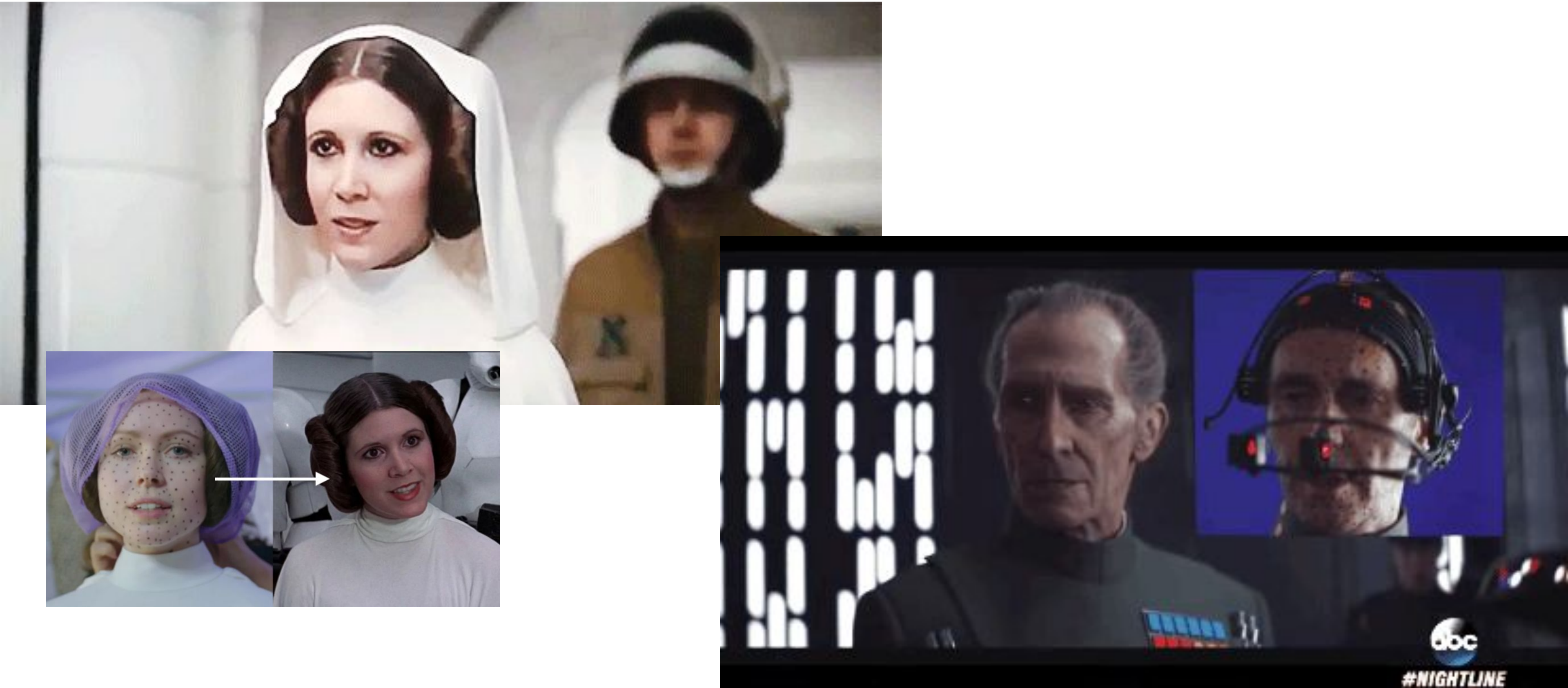
7 Best Face Swap Apps for Android and iOS (2018)

Kaushal September 28, 2018 Apps

TECHWISER

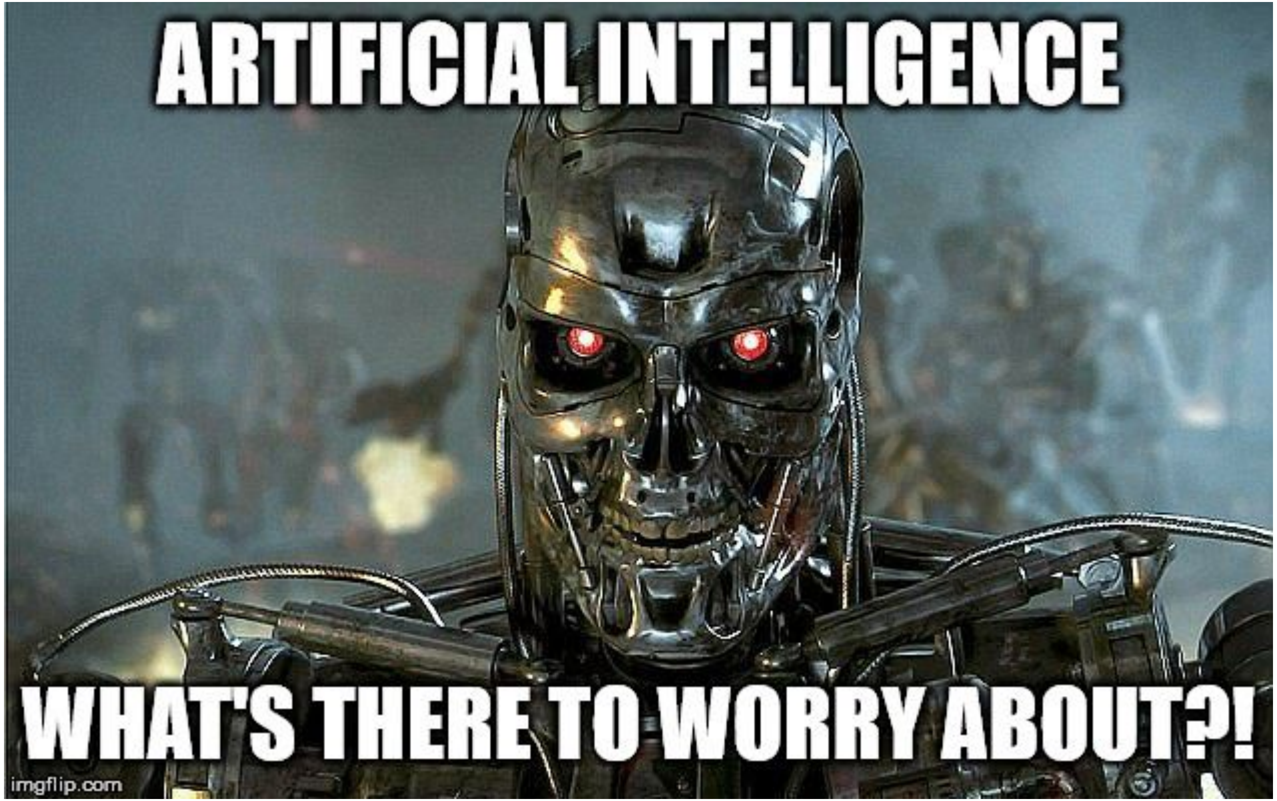
Nor is it your gramma’s video face swap

But I’m talking about cutting edge, Hollywood grade CGI:



Which could even be done **24 years ago**:

What’s new? The power of “Aaaa Iiiii”



Translation: anyone can create hyper realistic fake videos

And they can do it **at SCALE**

For example, the second (third?) coming of Nic Cage's career

More Nic Cage

And even this madness

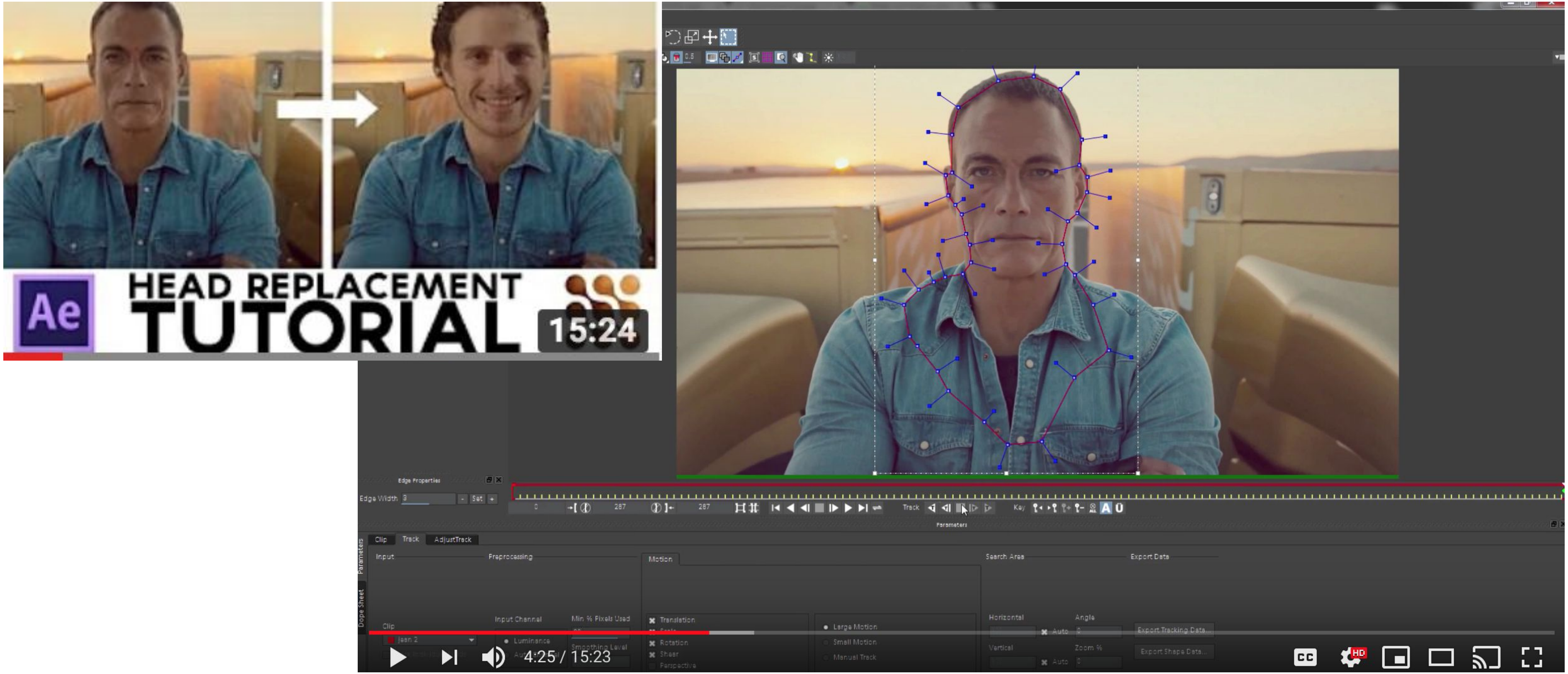
To summarize, “Deepfakes” refers to:

AI based tech to create hyper realistic videos with minimal skill, and at scale.

“AI generated synthetic media”

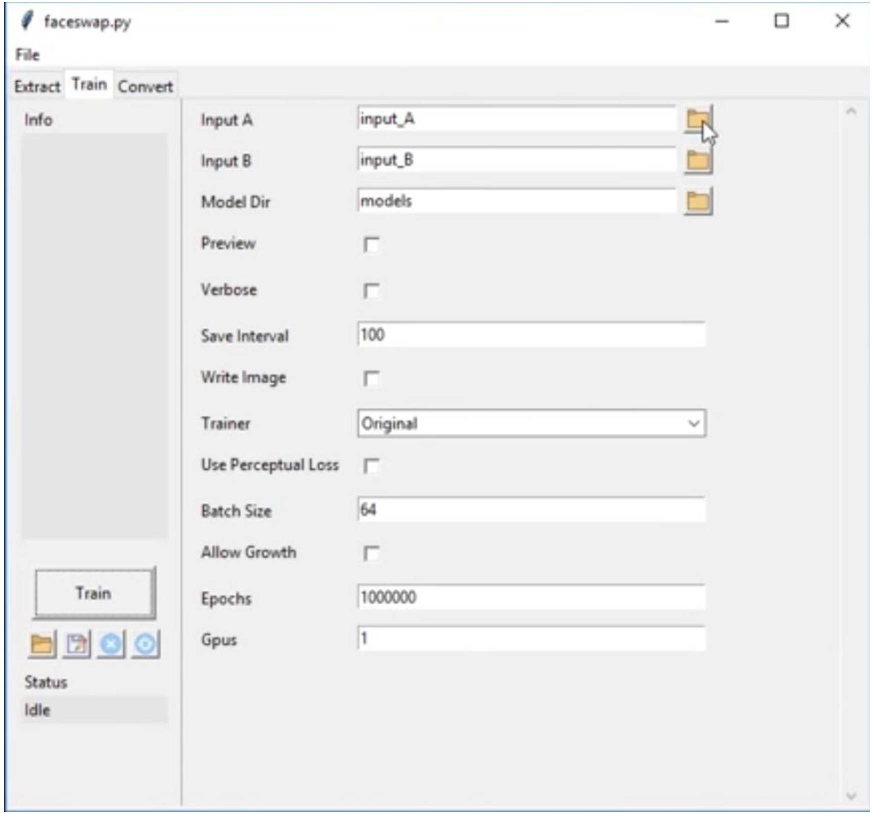
How do
they work?

How to make a video face swap the old way



How to Make a Deepfake - Step 1/3

Download and setup face swap app or code.



Idle

Input A: input_A

Input B: input_B

Model Dir: models

Preview: □

Verbose: □

Save Interval: 100

Write Image: □

Trainer: Original

Use Perceptual Loss: □

Batch Size: 64

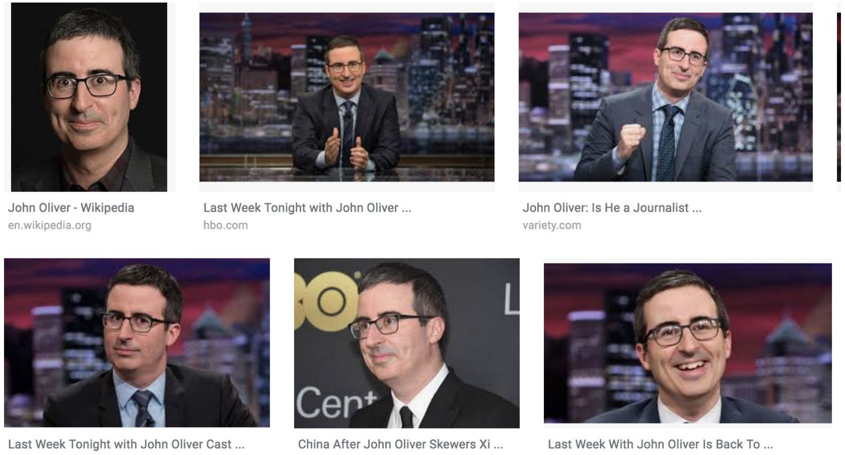
Allow Growth: □

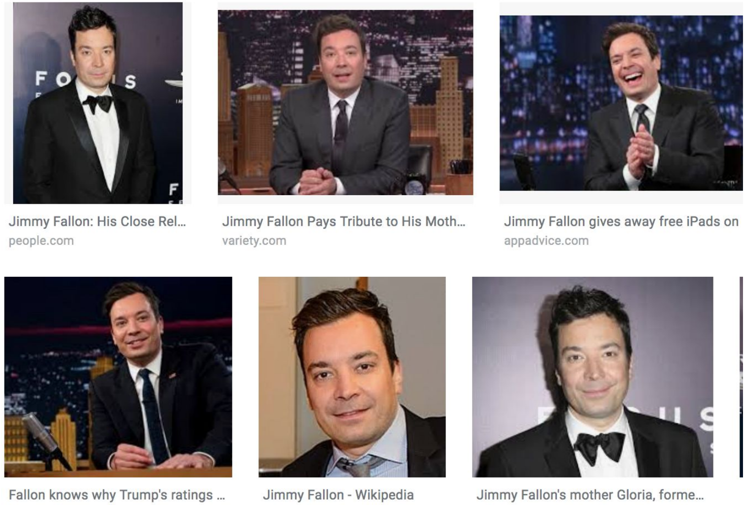
Epochs: 1000000

Gpus: 1

How to Make a Deepfake - Step 2/3

Get lots of photos and videos of both people.


John Oliver - Wikipedia
en.wikipedia.org
Last Week Tonight with John Oliver ...
hbo.com
John Oliver: Is He a Journalist ...
variety.com
Last Week Tonight with John Oliver Cast ...
China After John Oliver Skewers Xi ...
Last Week With John Oliver Is Back To ...

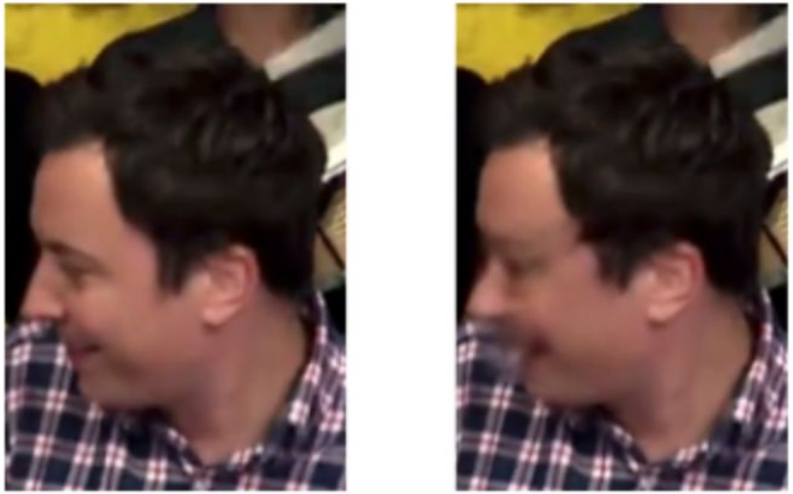

Jimmy Fallon: His Close Rel...
people.com
Jimmy Fallon Pays Tribute to His Moth...
variety.com
Jimmy Fallon gives away free iPads on
appadvice.com
Fallon knows why Trump's ratings ...
Jimmy Fallon - Wikipedia
Jimmy Fallon's mother Gloria, forme...

How to Make a Deepfake - Step 3/3

Click “Train” button and wait 24-72 hours.

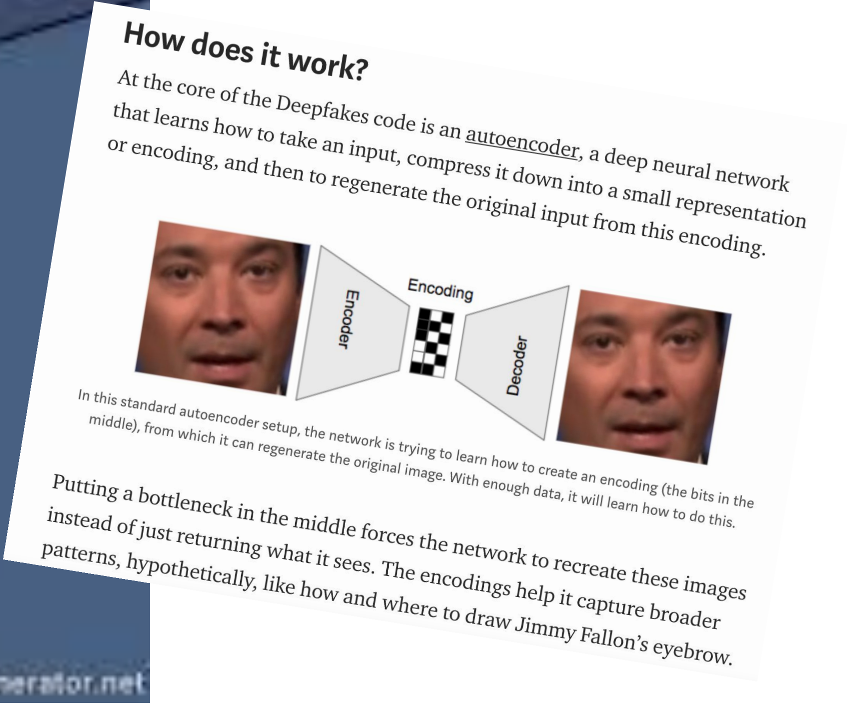
THAT’S IT!!

If things don’t work, get more data.



There just weren’t that many images of Oliver looking to the side, so the network was never able to learn by example and generate profile poses.

How does the algorithm work?



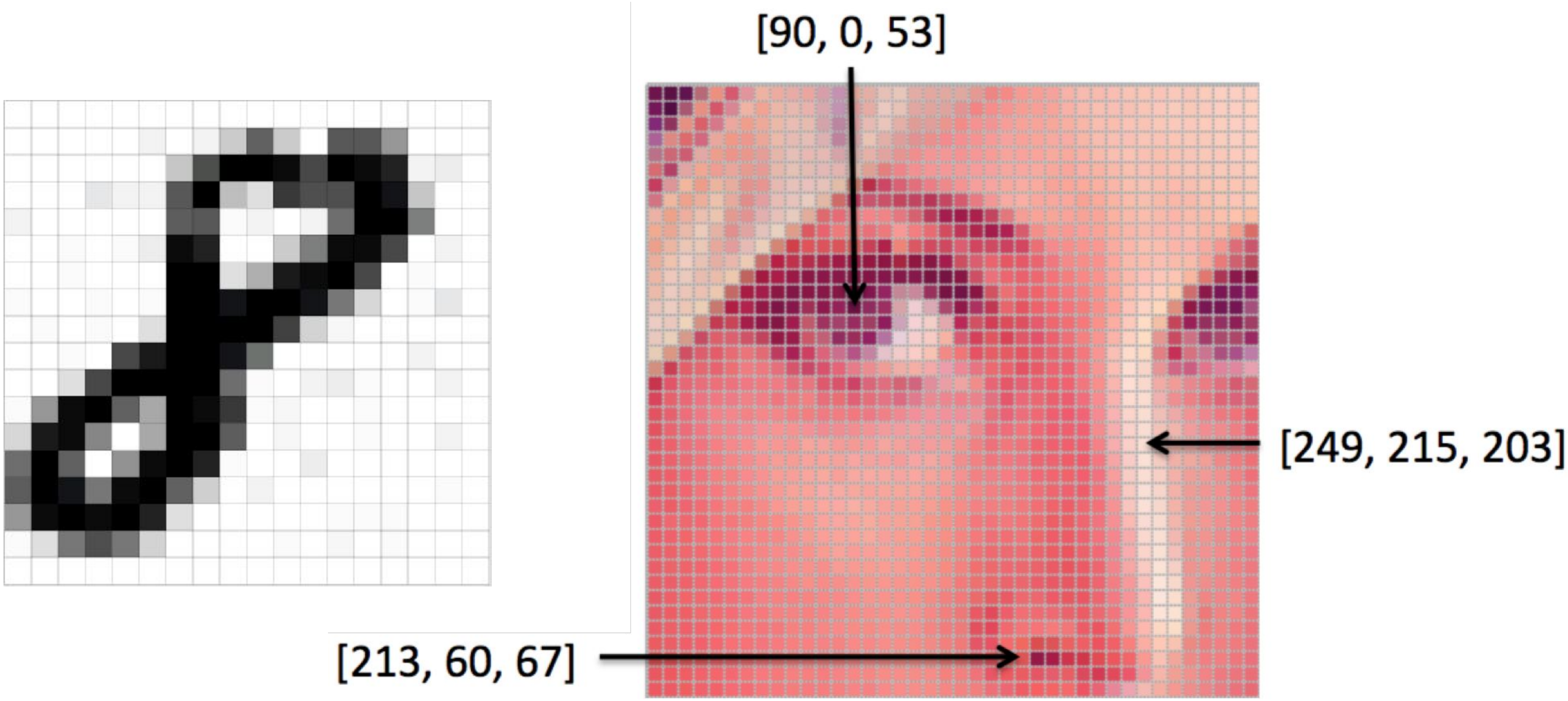
How does it work?

At the core of the Deepfakes code is an autoencoder, a deep neural network that learns how to take an input, compress it down into a small representation or encoding, and then to regenerate the original input from this encoding.

In this standard autoencoder setup, the network is trying to learn how to create an encoding (the bits in the middle), from which it can regenerate the original image. With enough data, it will learn how to do this.

Putting a bottleneck in the middle forces the network to recreate these images instead of just returning what it sees. The encodings help it capture broader patterns, hypothetically, like how and where to draw Jimmy Fallon's eyebrow.

First: an image can be represented as a grid of numbers



So generating an image, means generating a grid of numbers and displaying them

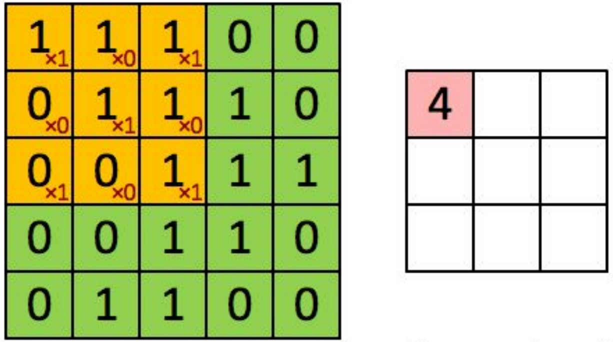
Manipulating an image, means adding/subtracting/multiplying/dividing those numbers

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

Image

1	0	1
0	1	0
1	0	1

Filter



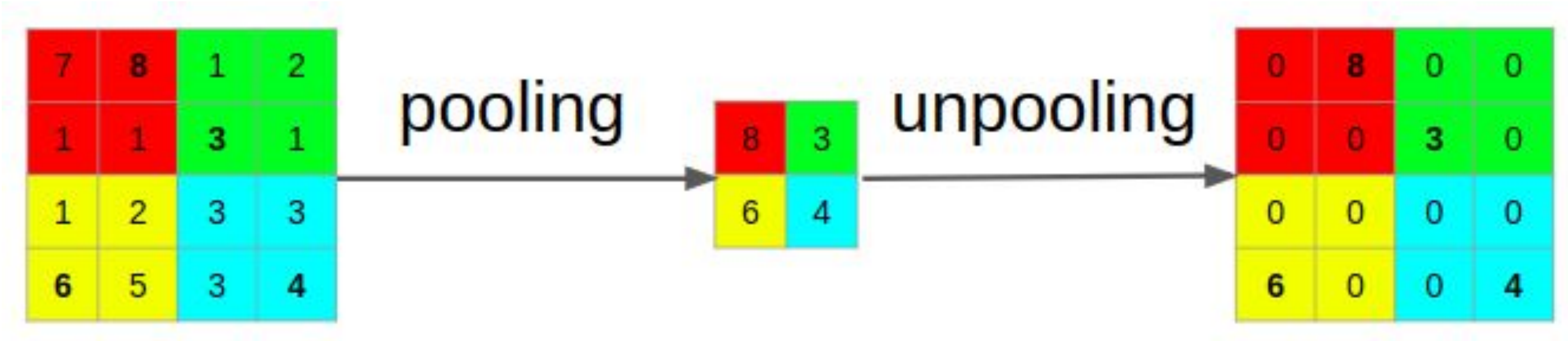
$1_{\times1}$	$1_{\times0}$	$1_{\times1}$	0	0
$0_{\times0}$	$1_{\times1}$	$1_{\times0}$	1	0
$0_{\times1}$	$0_{\times0}$	$1_{\times1}$	1	1
0	0	1	1	0
0	1	1	0	0

Image

4		

Convolved Feature

Images can be shrunk by throwing away information, or expanded by generating it

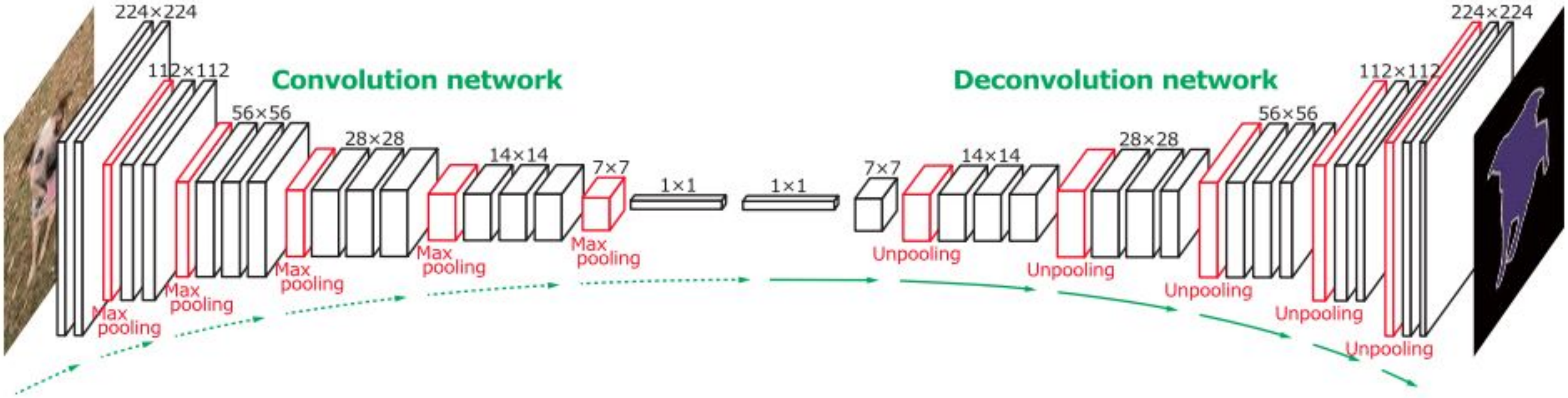


7	**8**	1	2
1	1	**3**	1
1	2	3	3
6	5	3	**4**

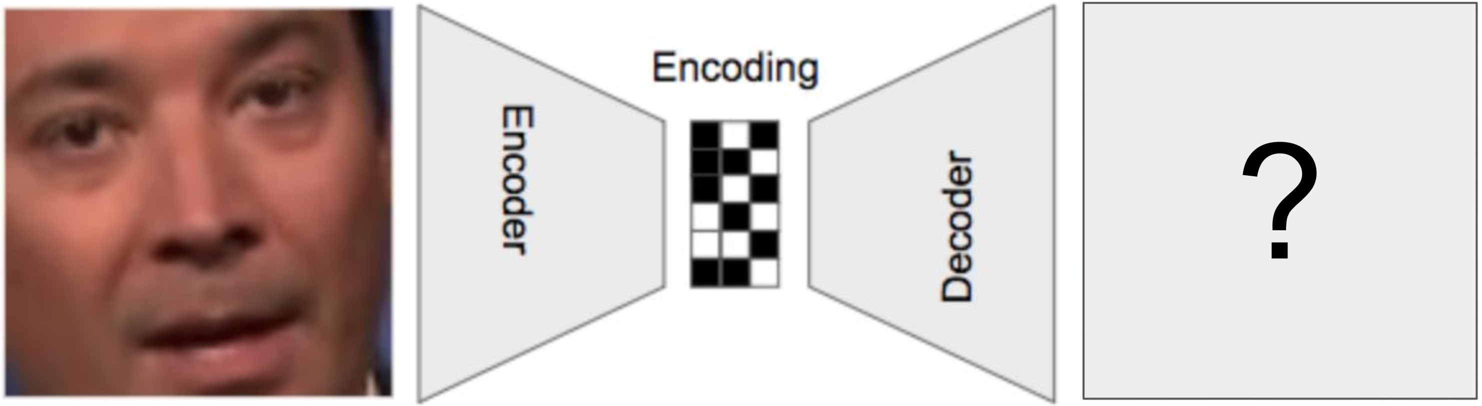
8	3
6	4

0	**8**	0	0
0	0	**3**	0
0	0	0	0
6	0	0	**4**

Stack these together to create a beast of a neural network with millions of parameters



Capture (encode) the essence of the image



In this standard autoencoder setup, the network is trying to learn how to create an encoding (the bits in the middle), from which it can regenerate the original image. With enough data, it will learn how to do this.

Train the network:compare it's output with what was expected,

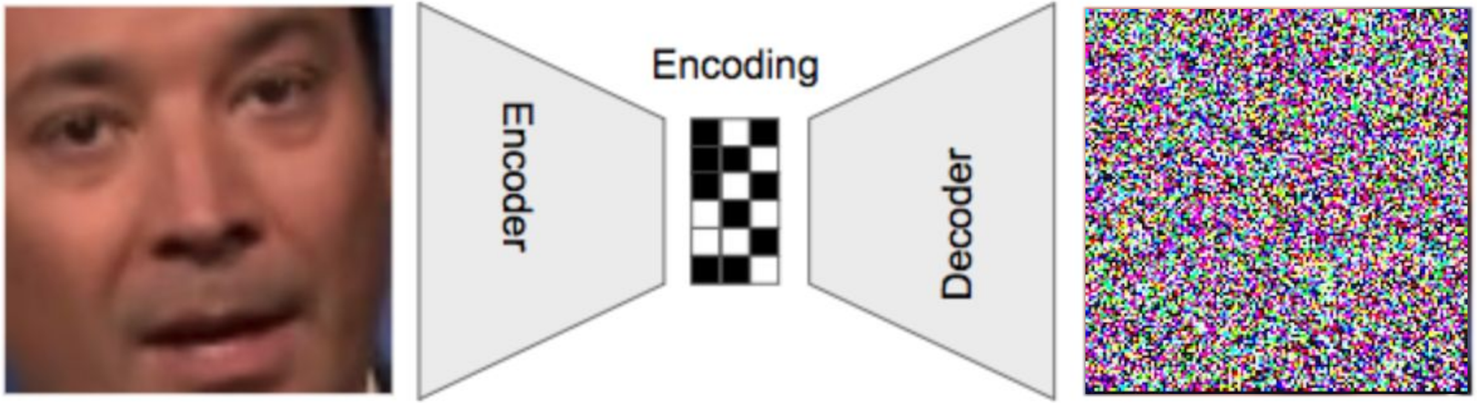
Wanted this

Got this

Calculate the difference between expected and predicted, and adjust weights.

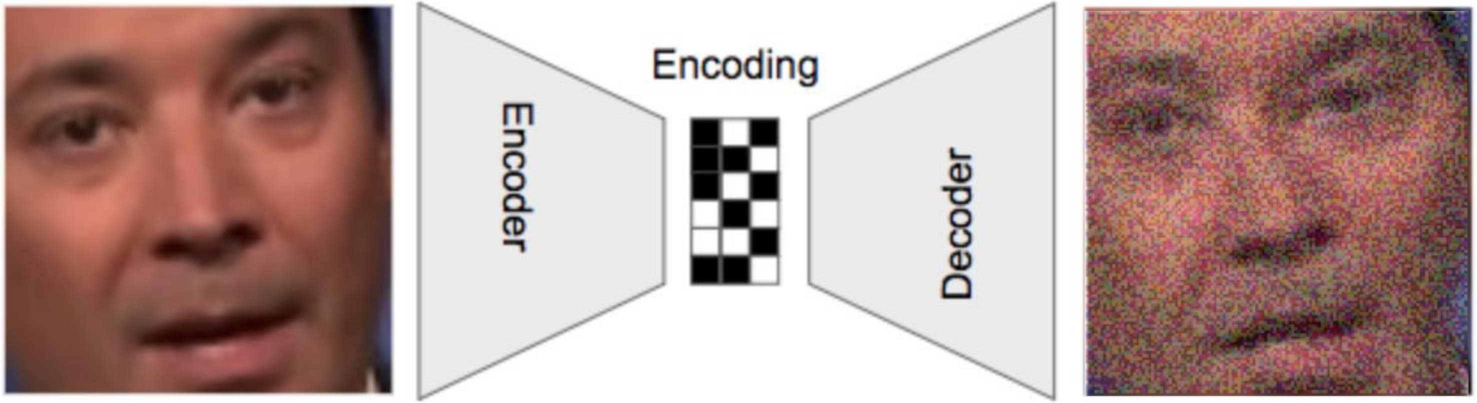
This is "training"

Training means comparing the output, with desired, and trying again



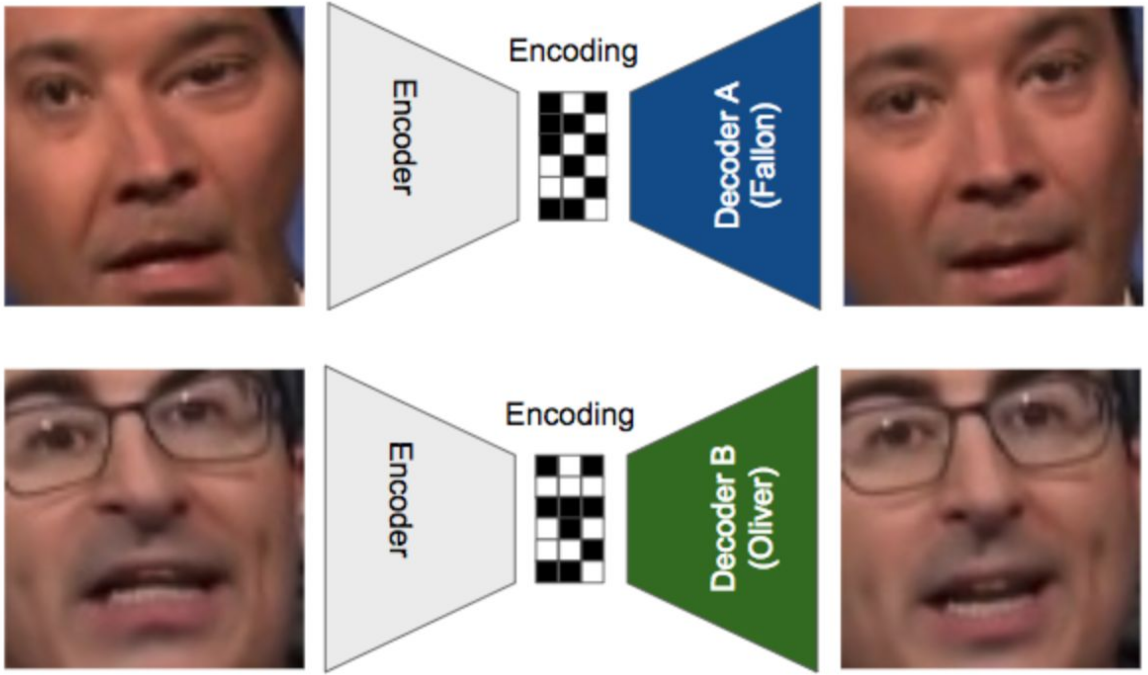
In this standard autoencoder setup, the network is trying to learn how to create an encoding (the bits in the middle), from which it can regenerate the original image. With enough data, it will learn how to do this.

After many iterations, it will eventually learn the right “parameters” to tweak the input into the desired output



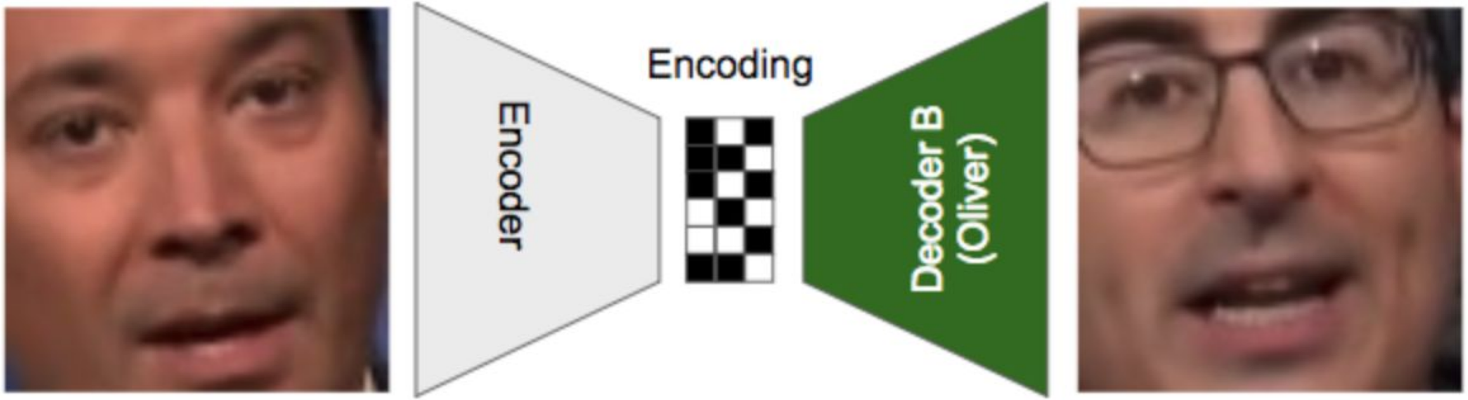
In this standard autoencoder setup, the network is trying to learn how to create an encoding (the bits in the middle), from which it can regenerate the original image. With enough data, it will learn how to do this.

Deepfakes: Train 1 encoder, 2 decoders



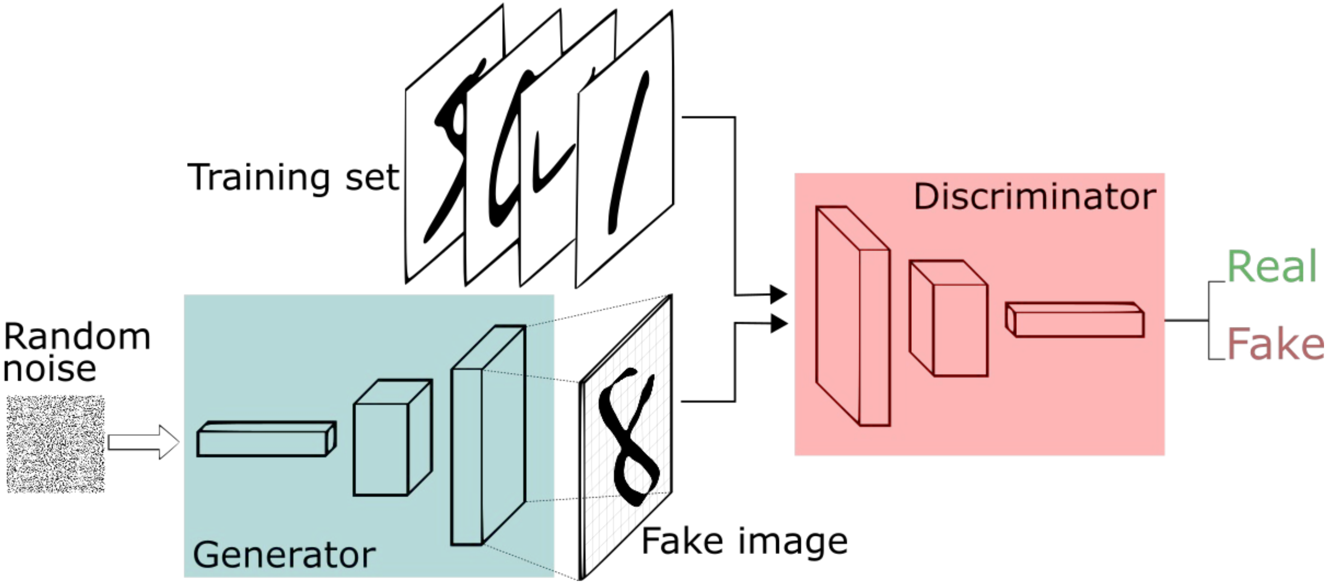
There is only one encoder that is shared between the Fallon and Oliver cases, but the decoders are different. During training, the input faces are warped, to simulate the notion of "we want a face kind of like this".

Finally, swap the decoders and voila!

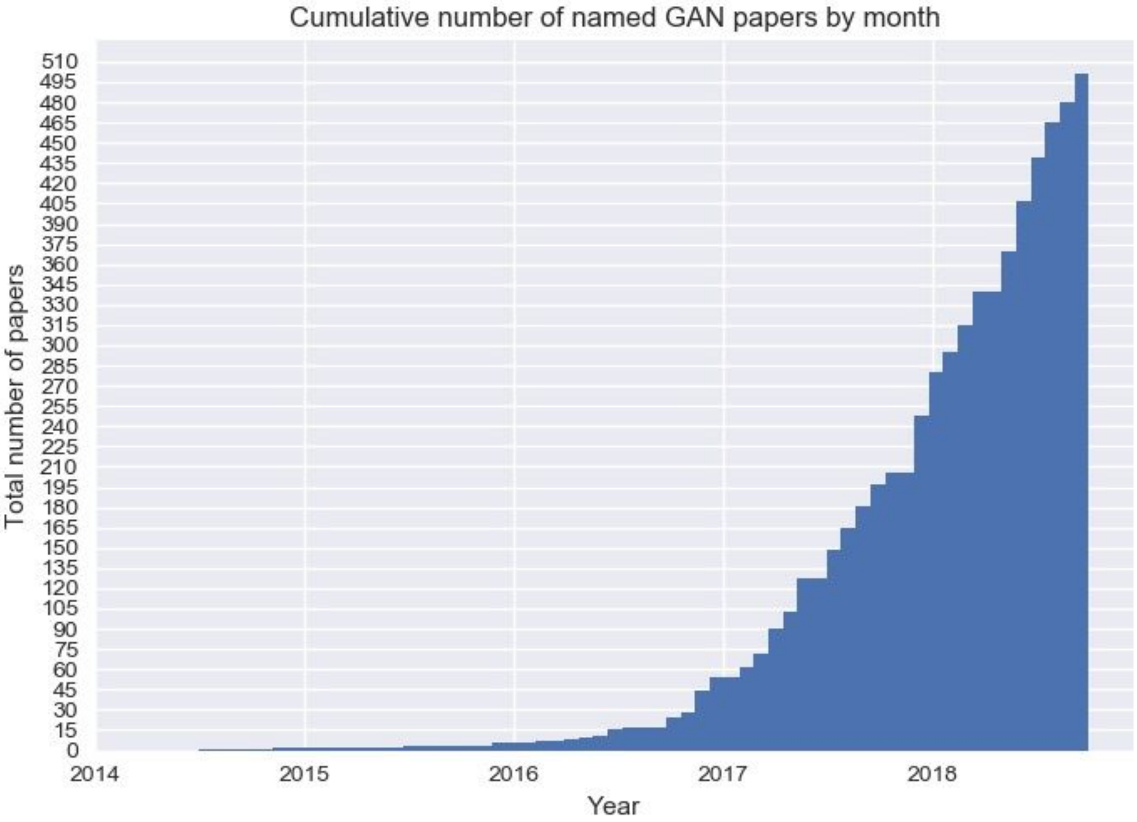


This is how we run the model to generate images. The encoder captures the essence of Fallon's face and gives it to Decoder B, which says "ah, another noisy input, but I've learned how to turn this into Oliver... voila!"

That was an *autoencoder*.
More popular now is the
generative adversarial network, or GAN.



GANs are a hot research area



Cumulative number of named GAN papers by month

Ian Goodfellow

@goodfellow_ian

Google Brain research scientist leading a team studying adversarial techniques in AI. Lead author of deeplearningbook.org

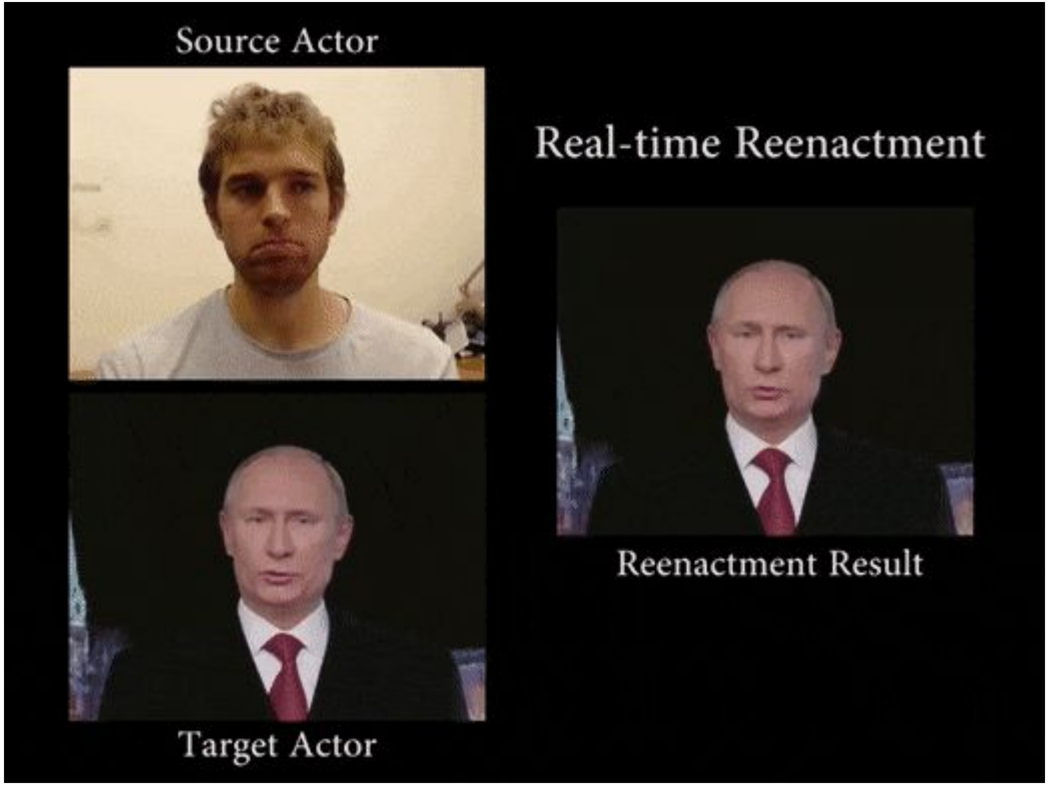
San Francisco, CA

iangoodfellow.com

Joined September 2016

What can Deepfakes do today?

Lip Syncing


Source Actor
Real-time Reenactment
Reenactment Result
Target Actor

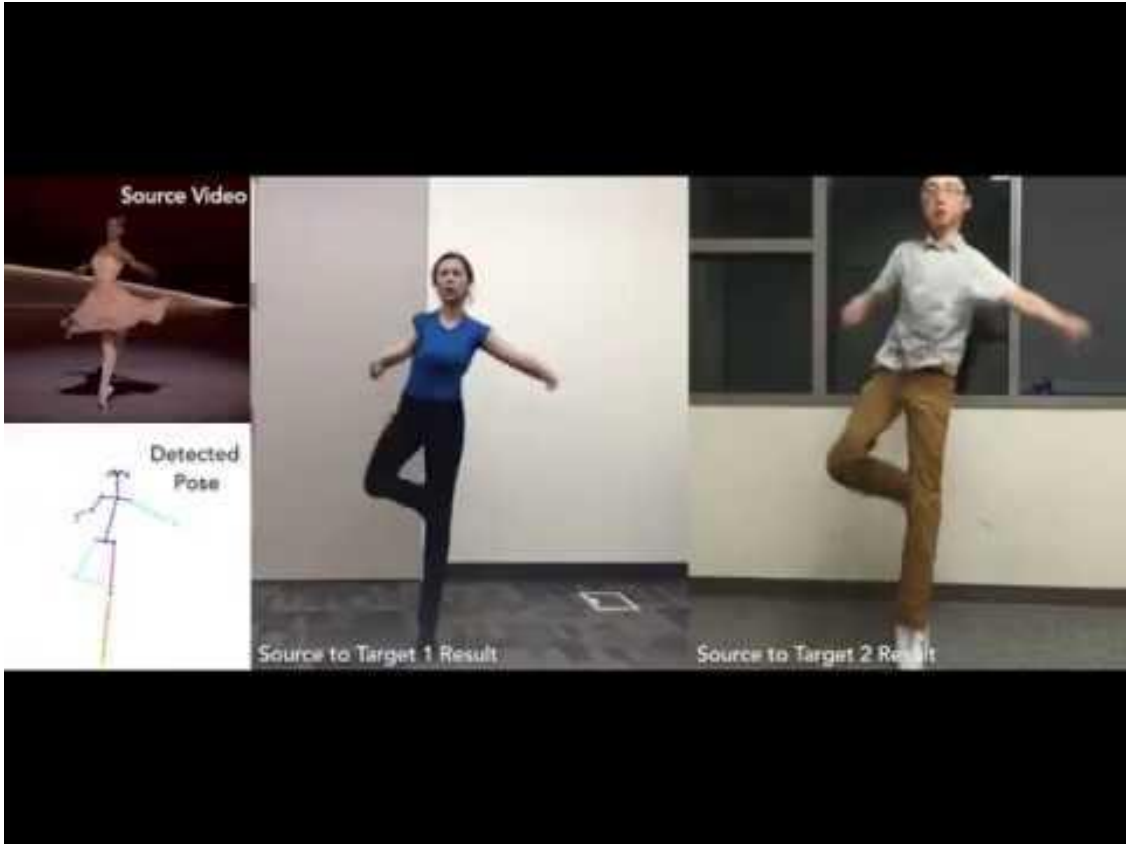
Live Puppeting

H. Kim [1] P. Garrido [1] A. Tewari [1] W. Xu [1] J. Thies [2] M. Nießner [2] P. Perez [3] C. Richardt [4] M. Zollhöfer [5] C. Theobalt [1]

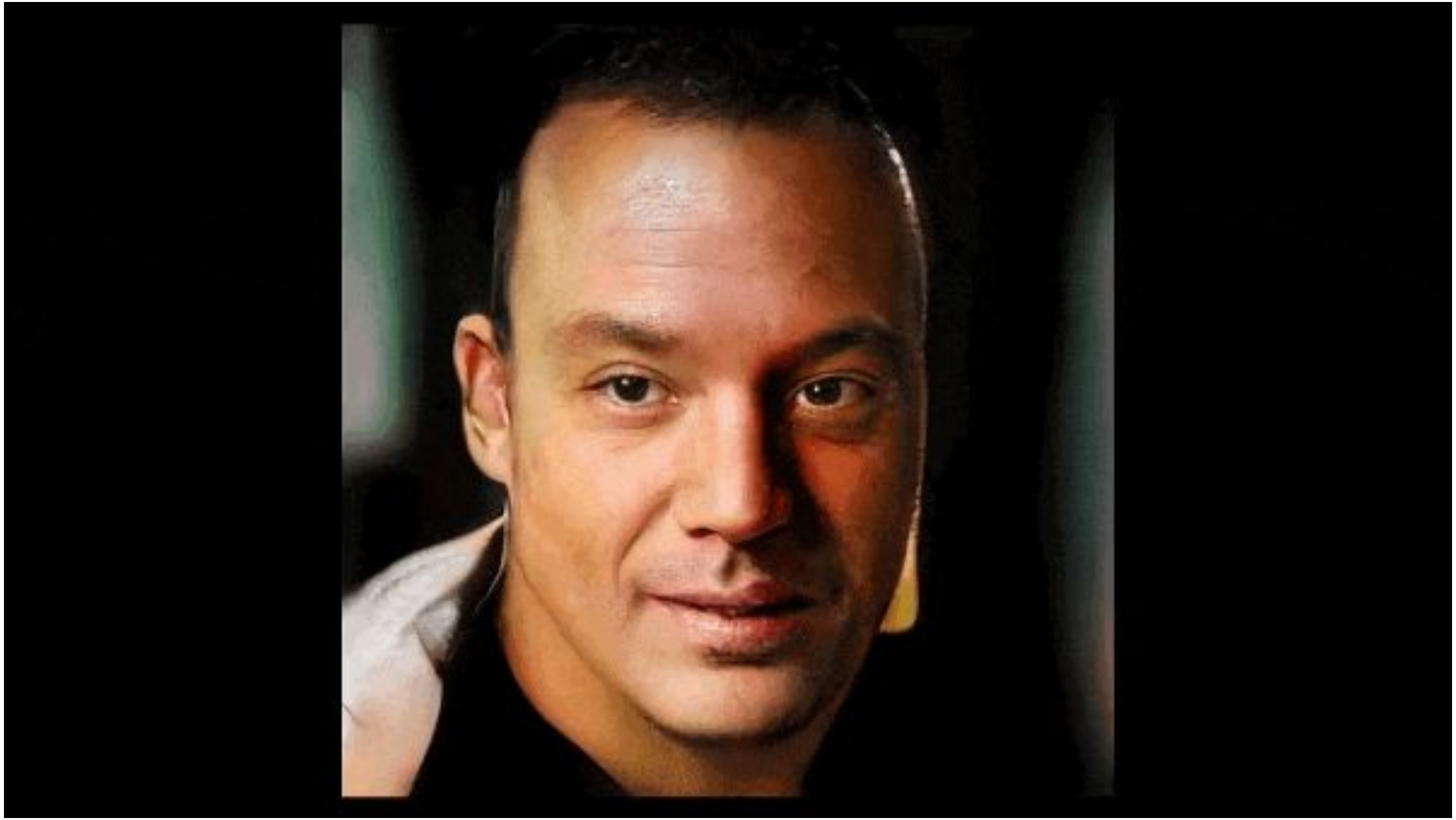
[1] MPI Informatics [2] Technical University of Munich [3] Technicolor [4] University of Bath [5] Stanford University

Impact of Audio + Video is huge

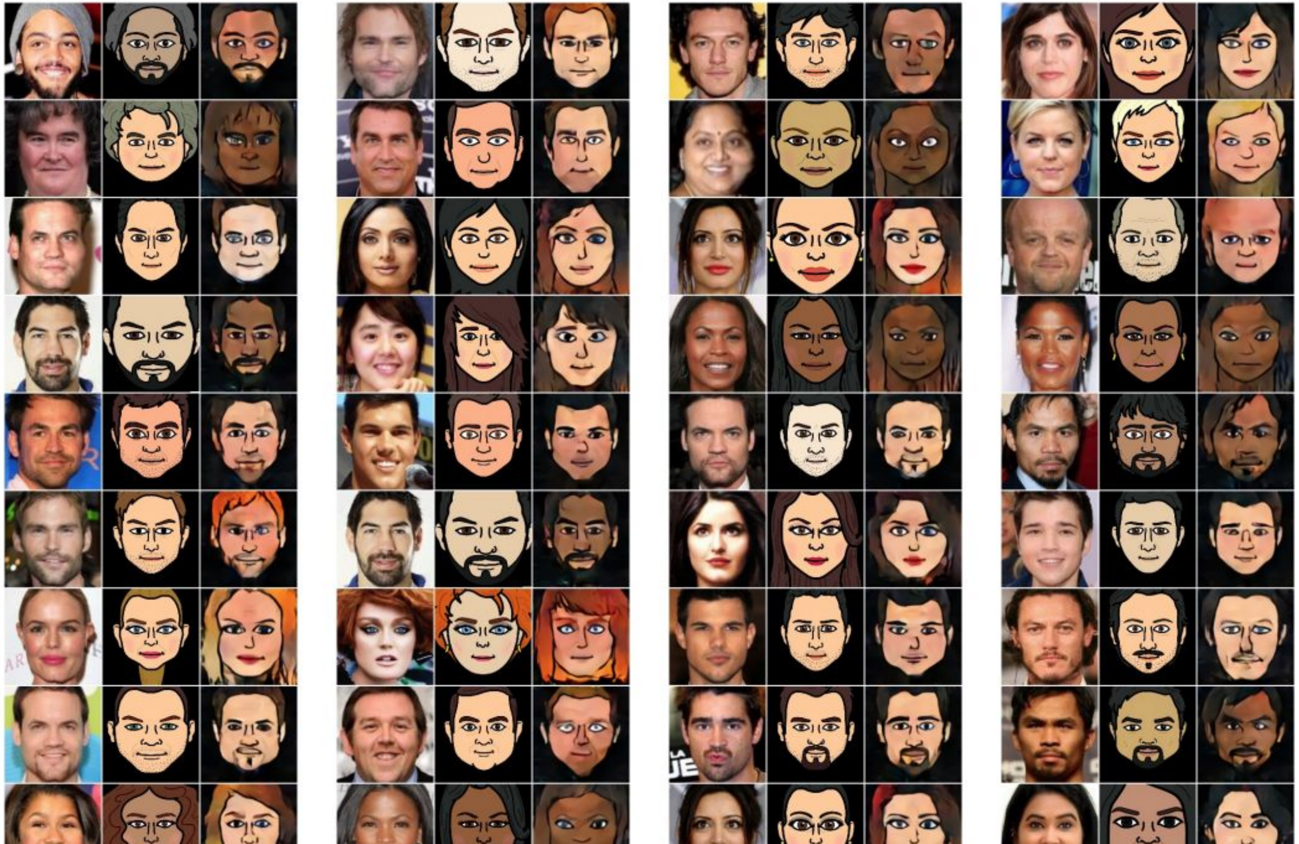
Dancing


Source Video
Detected Pose
Source to Target 1 Result
Source to Target 2 Result

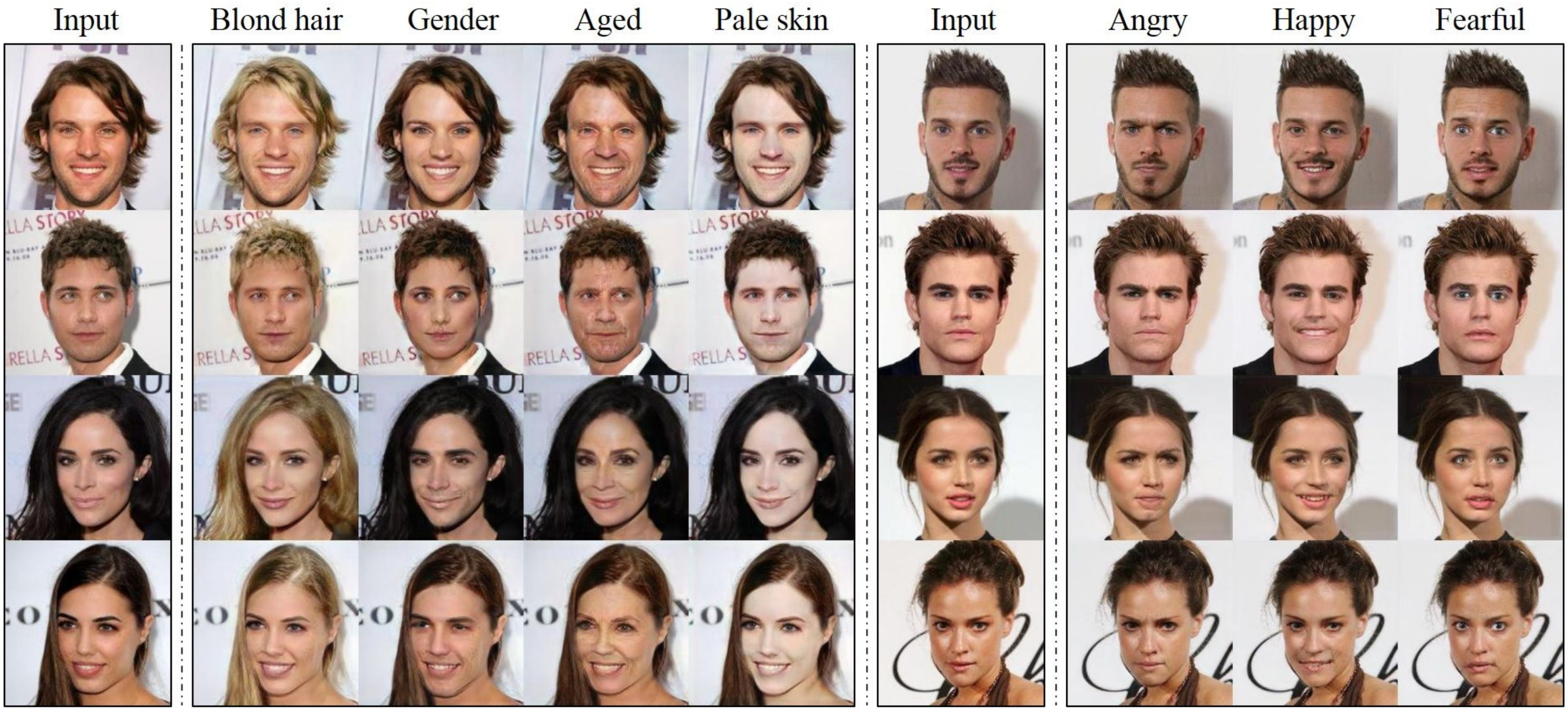
Real Celebrity Face Generation



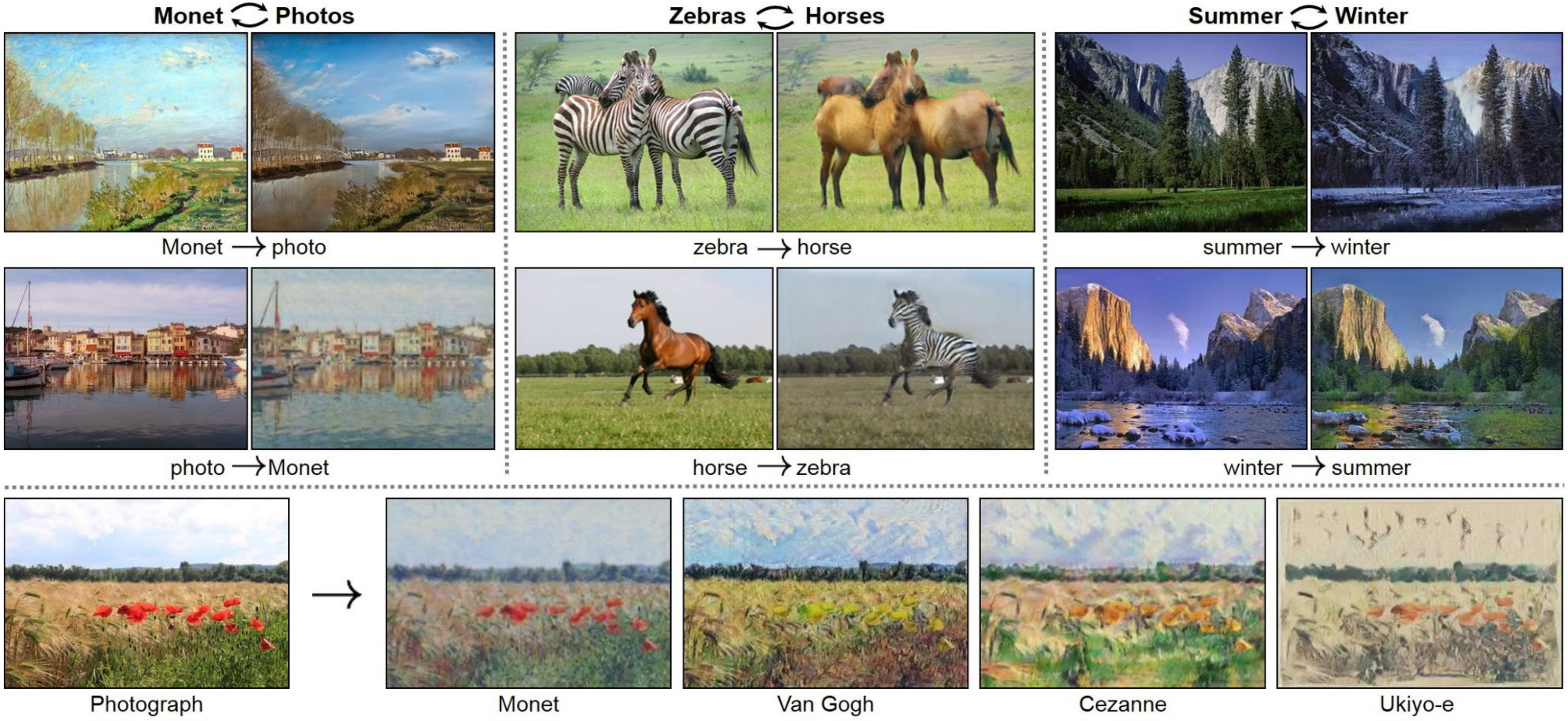
Face to Emoji



Changing Emotion, Hair, Age

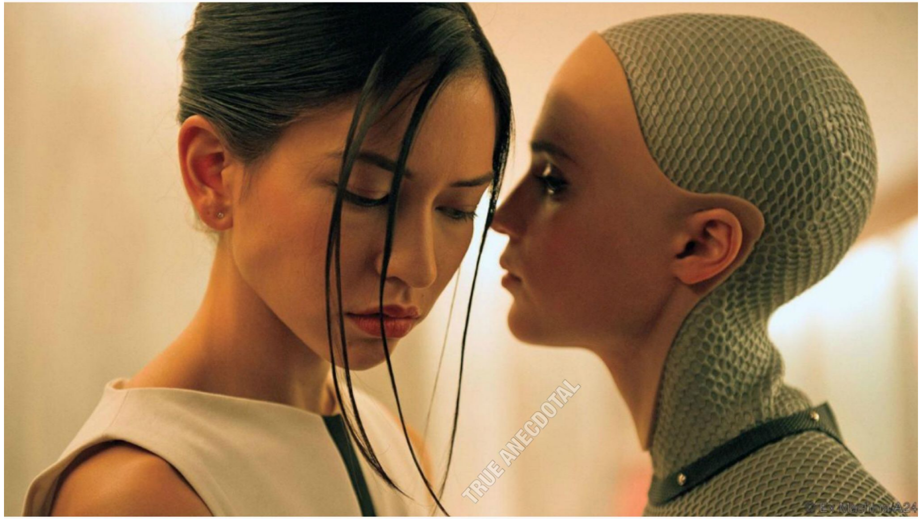


Transforming Summer to Winter and Day to Night



Generating life like audio

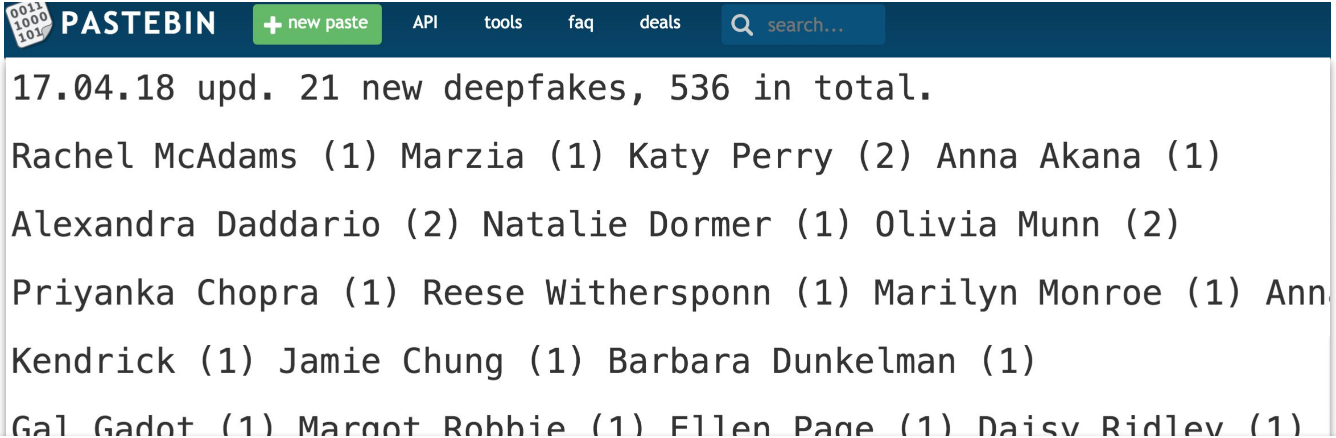
GOOGLE DUPLEX: “Hi, I’d like to book a women’s haircut for a client.”



GOOGLE DUPLEX: “Also, could you go ahead and kill my client while I escape the Googleplex?”

Aren’t Deepfakes going to be terrible for society?

Synthetic porn for sale on pastebin



17.04.18 upd. 21 new deepfakes, 536 in total.

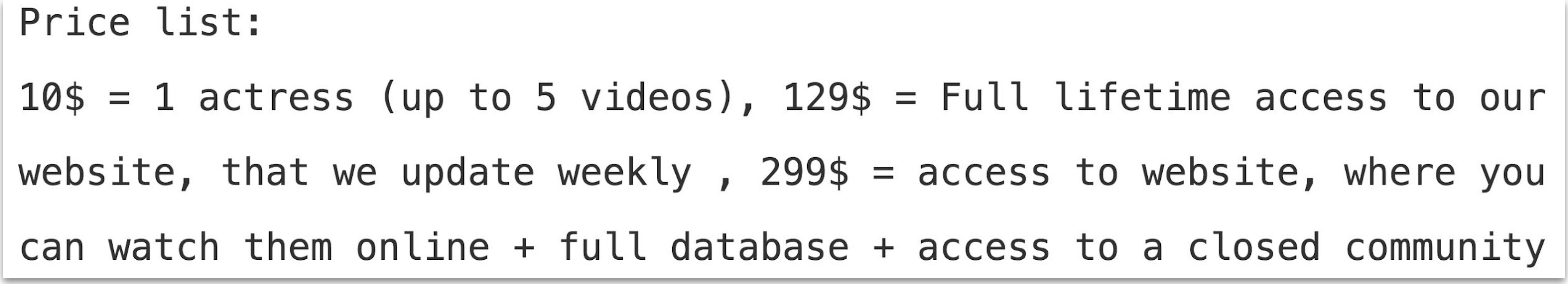
Rachel McAdams (1) Marzia (1) Katy Perry (2) Anna Akana (1)

Alexandra Daddario (2) Natalie Dormer (1) Olivia Munn (2)

Priyanka Chopra (1) Reese Withersponn (1) Marilyn Monroe (1) Ann

Kendrick (1) Jamie Chung (1) Barbara Dunkelman (1)

Gal Gadot (1) Margot Robbie (1) Ellen Page (1) Daisy Ridley (1)

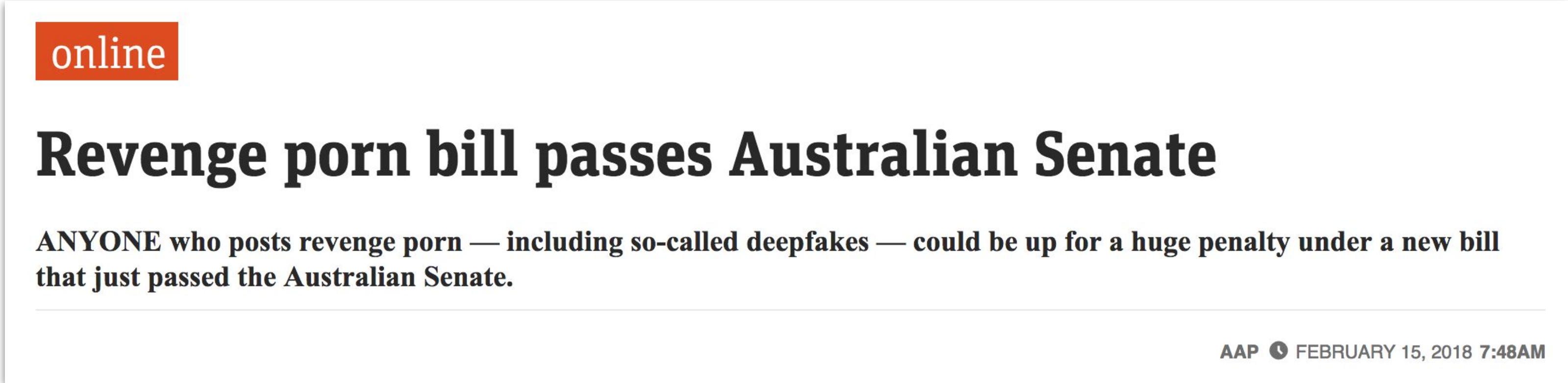


Price list:

10$ = 1 actress (up to 5 videos), 129$ = Full lifetime access to our website, that we update weekly , 299$ = access to website, where you can watch them online + full database + access to a closed community

Sextortion

Btw, she changed the law. What a hero!



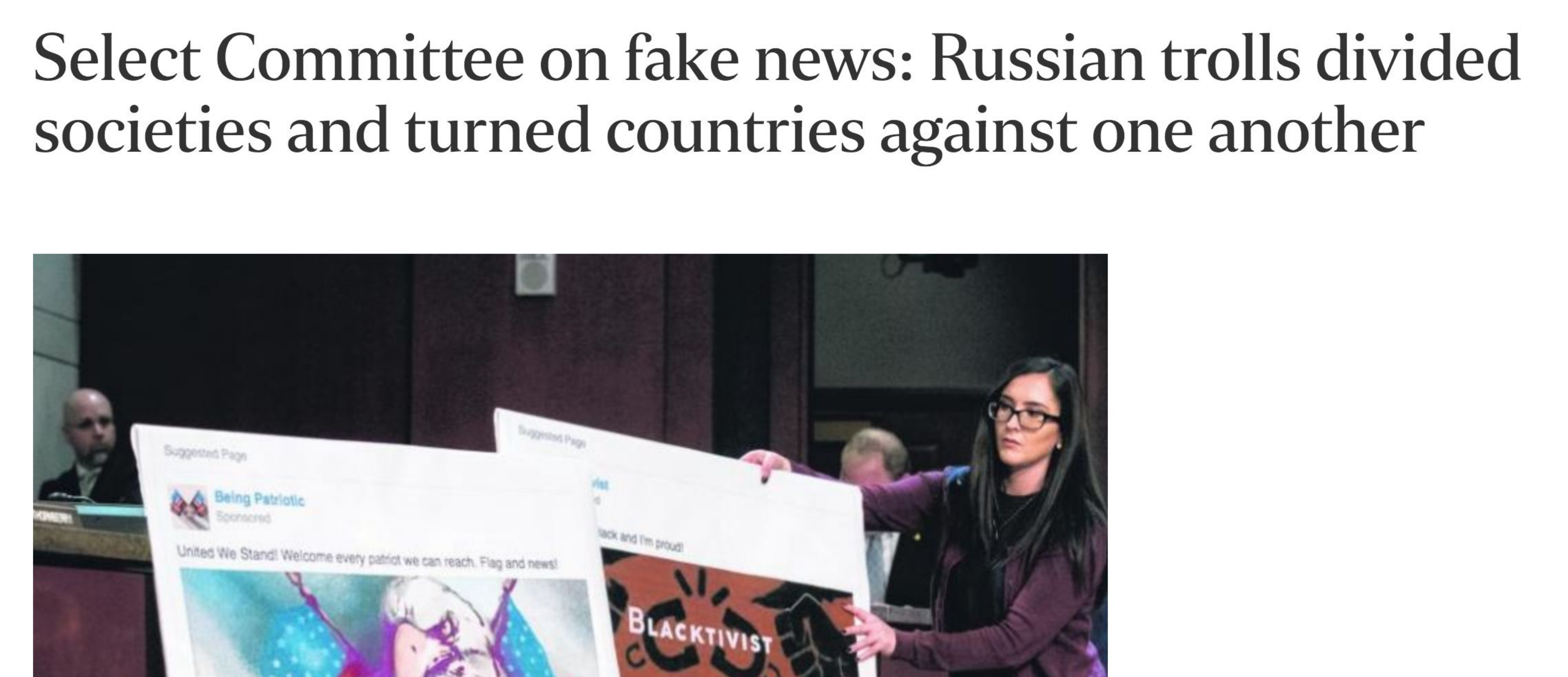
online

Revenge porn bill passes Australian Senate

ANYONE who posts revenge porn — including so-called deepfakes — could be up for a huge penalty under a new bill that just passed the Australian Senate.

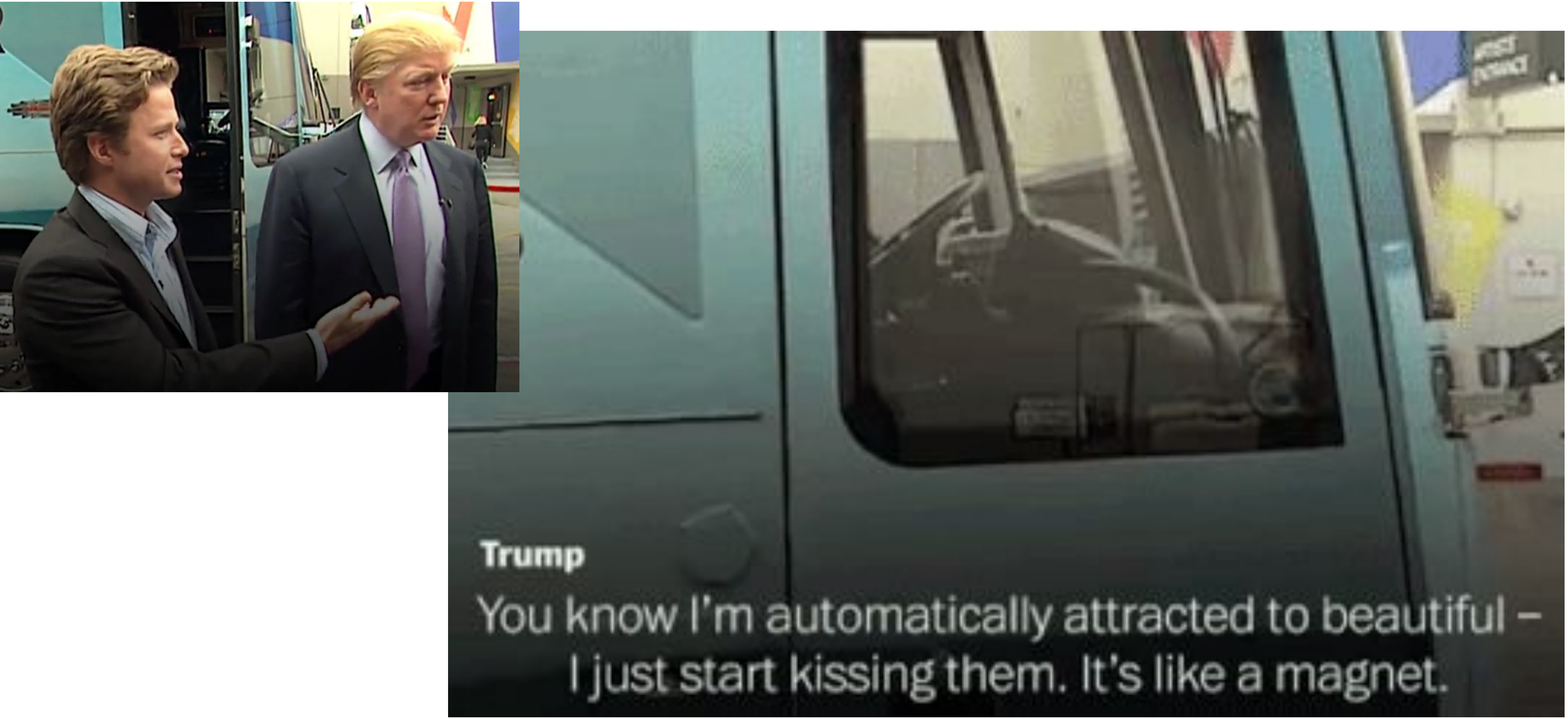
AAP FEBRUARY 15, 2018 7:48AM

Political instability



Select Committee on fake news: Russian trolls divided societies and turned countries against one another

Denying real video



Trump

You know I'm automatically attracted to beautiful –
I just start kissing them. It's like a magnet.

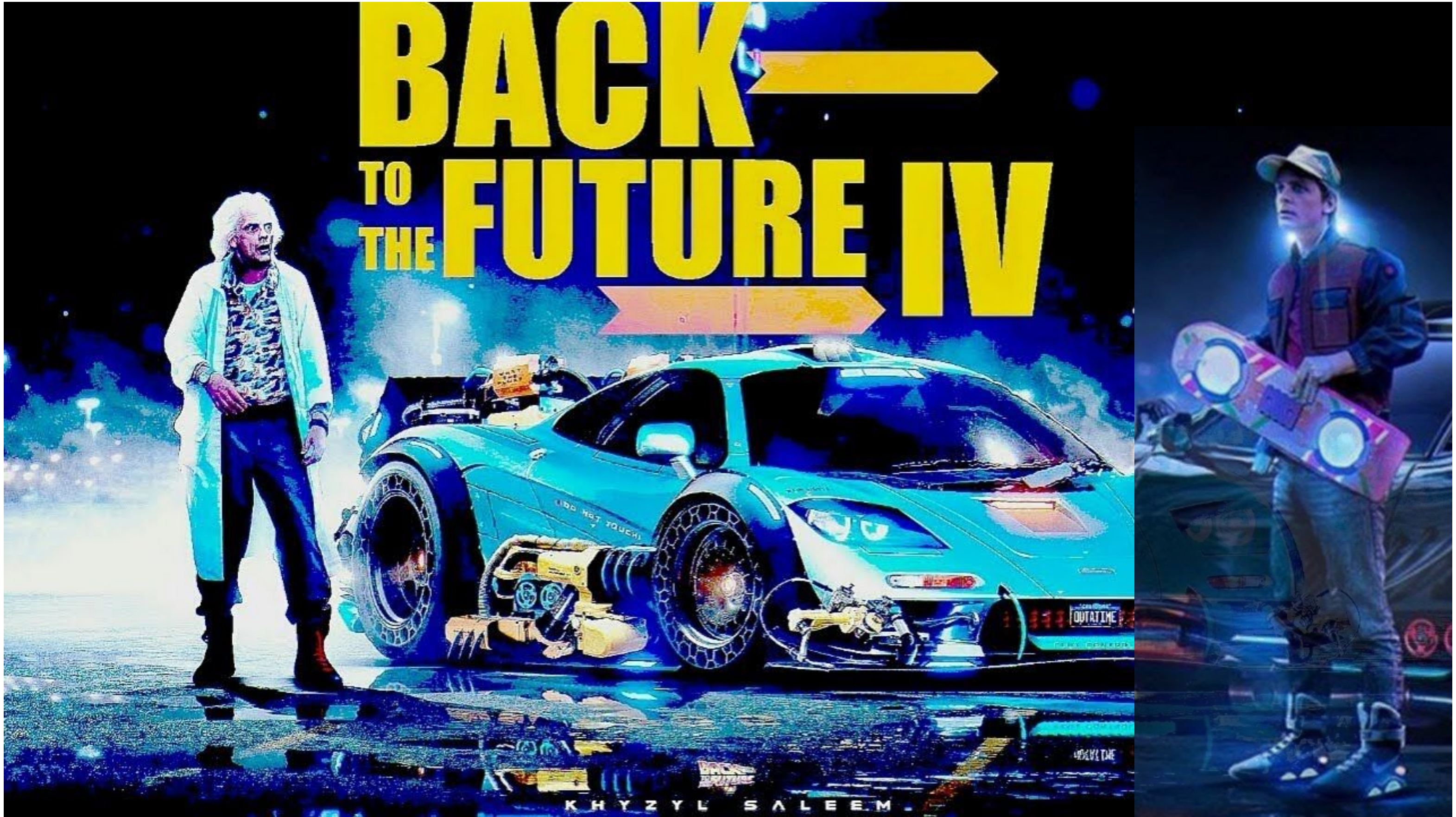
Jeopardizing witness video authenticity

WITNESS makes it possible for anyone, anywhere to use video and technology to protect and defend human rights.

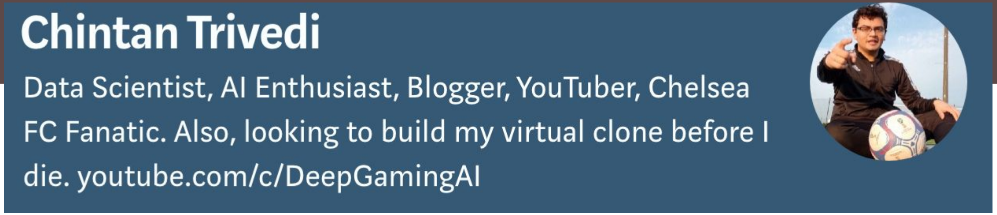
Are there any constructive uses for Deepfakes?

Special effects: Leia by “Derp Fakes”

Bringing back old favorites



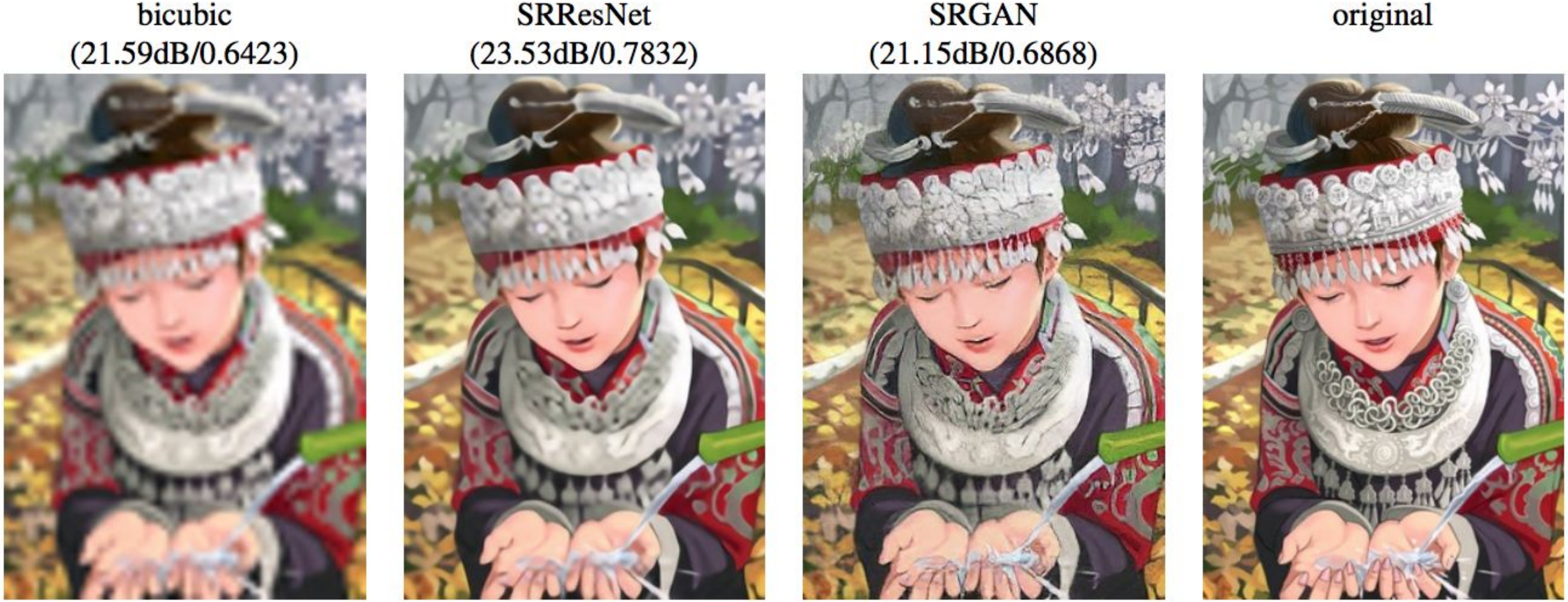
Enhancing video games



Chintan Trivedi

Data Scientist, AI Enthusiast, Blogger, YouTuber, Chelsea FC Fanatic. Also, looking to build my virtual clone before I die. youtube.com/c/DeepGamingAI

Enhancing low res images



bicubic
(21.59dB/0.6423)

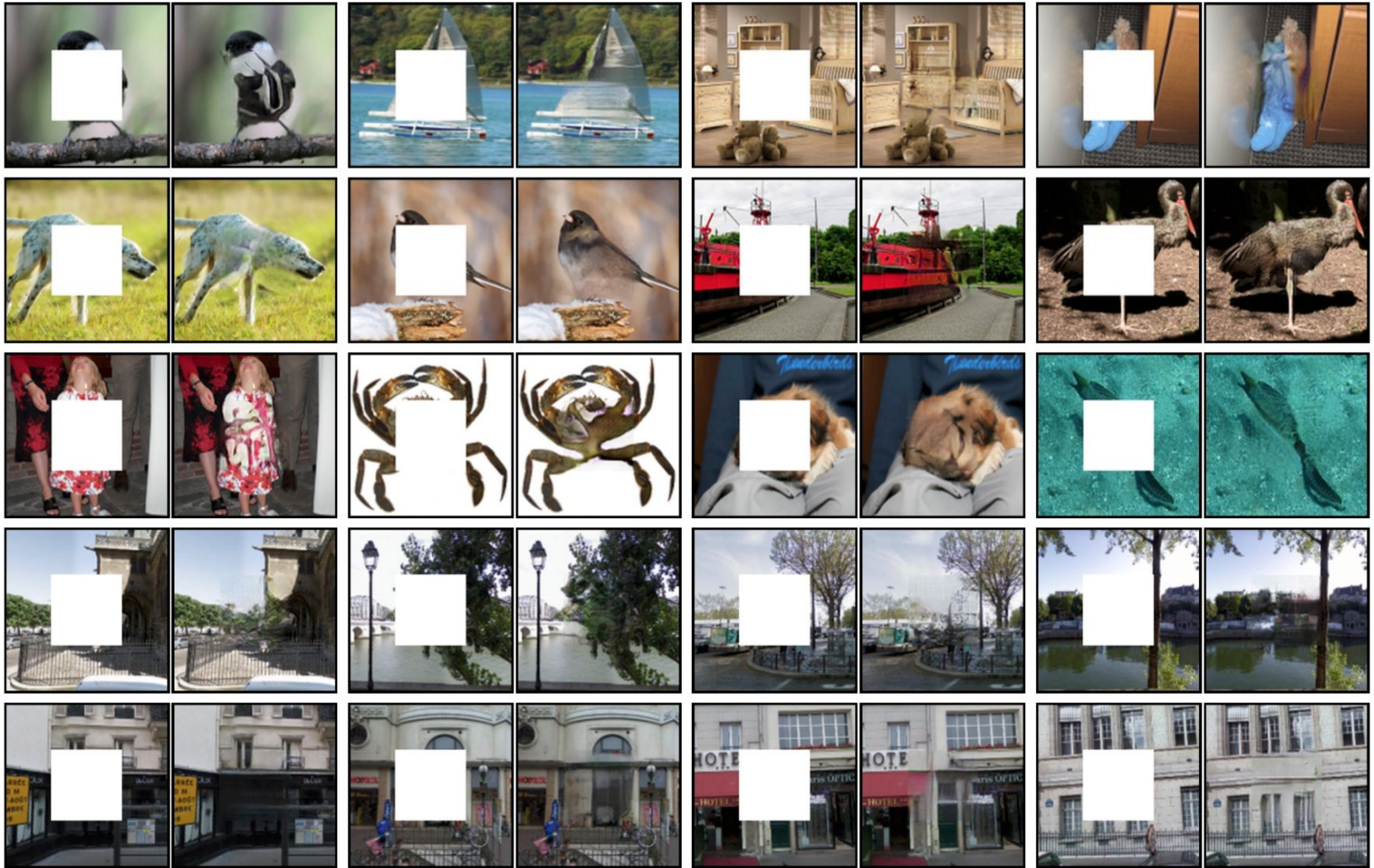
SRResNet
(23.53dB/0.7832)

SRGAN
(21.15dB/0.6868)

original

Figure 2: From left to right: bicubic interpolation, deep residual network optimized for MSE, deep residual generative adversarial network optimized for a loss more sensitive to human perception, original HR image. Corresponding PSNR and SSIM are shown in brackets. [4× upscaling]

Filling in gaps



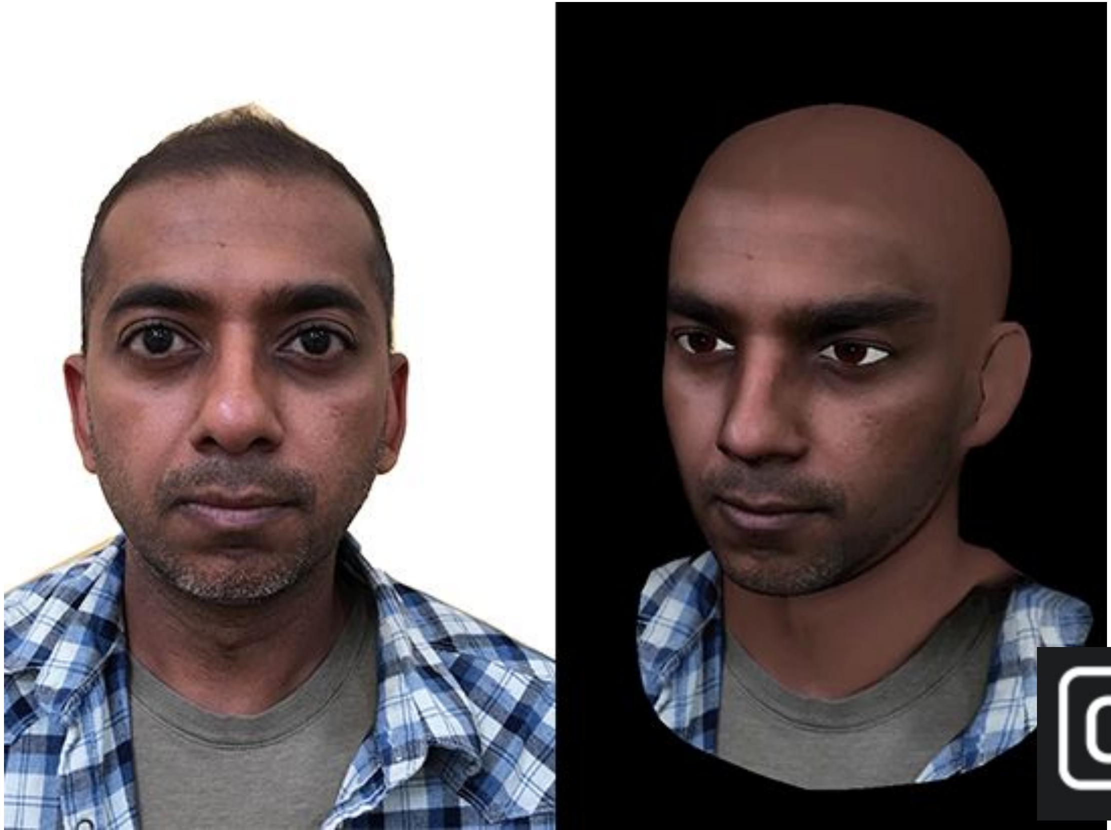
Education: legendary physics lectures by Feynman himself?

It’s not too late for Neil Degrasse Tyson!

Interact with history or deceased loved ones

USC Institute for Creative Technologies

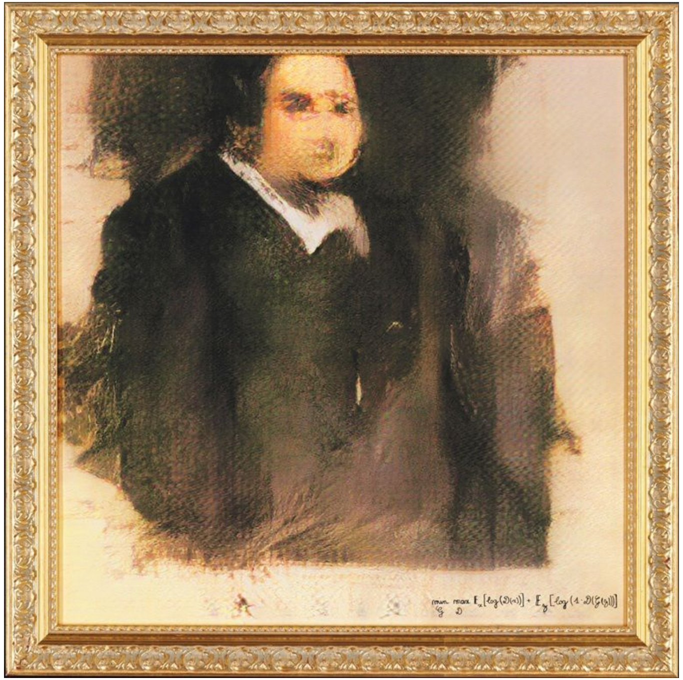
Create digital avatars for VR/AR applications (1 selfie → avatar!)



ObEN

ART

AI Portrait fetches $432,500 (43x estimate)


$\min_{G} \max_{D} \mathbb{E}_x[\log(D(x))] + \mathbb{E}_z[\log(1-D(G(z)))]$

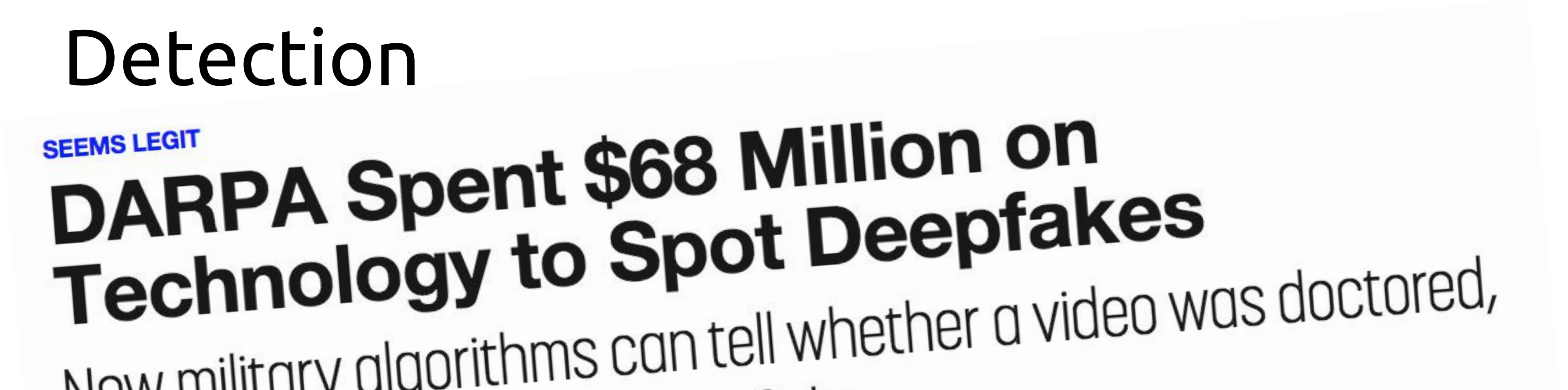
What should we do to prepare for a post Deepfakes world?

Educate: raise awareness.

Regulate: applications, not research

Making viruses is not illegal. Distributing them is.

Detection



SEEMS LEGIT

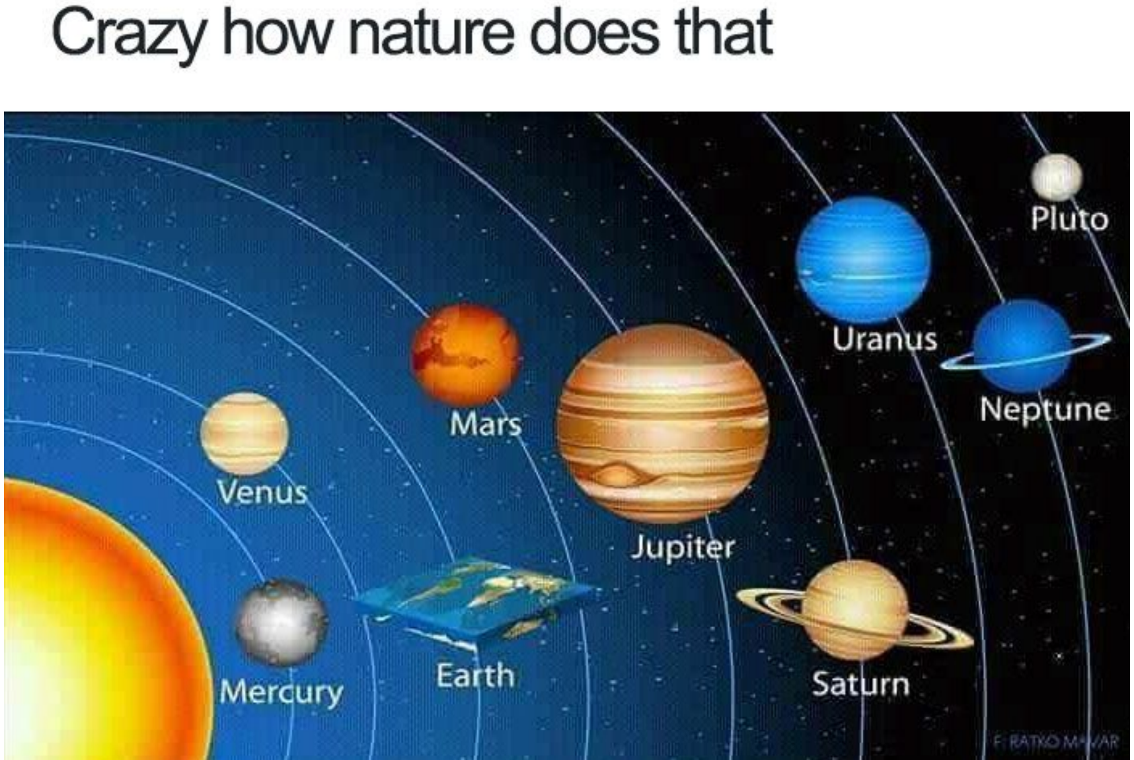
DARPA Spent $68 Million on Technology to Spot Deepfakes

New military algorithms can tell whether a video was doctored, but DARPA think it's losing the fight.

Dan Robitzski | November 19th 2018

Dr. Hany Farid
UC Berkeley

Distribution: the most clickbaity thing



Authenticity from Source

Thank you!

For more, Google: “Exploring Deepfakes”

Gaurav Oberoi
goberoi@twitter/gmail/medium