

AI for the Media

Week 9, DeepFakes



Motivation

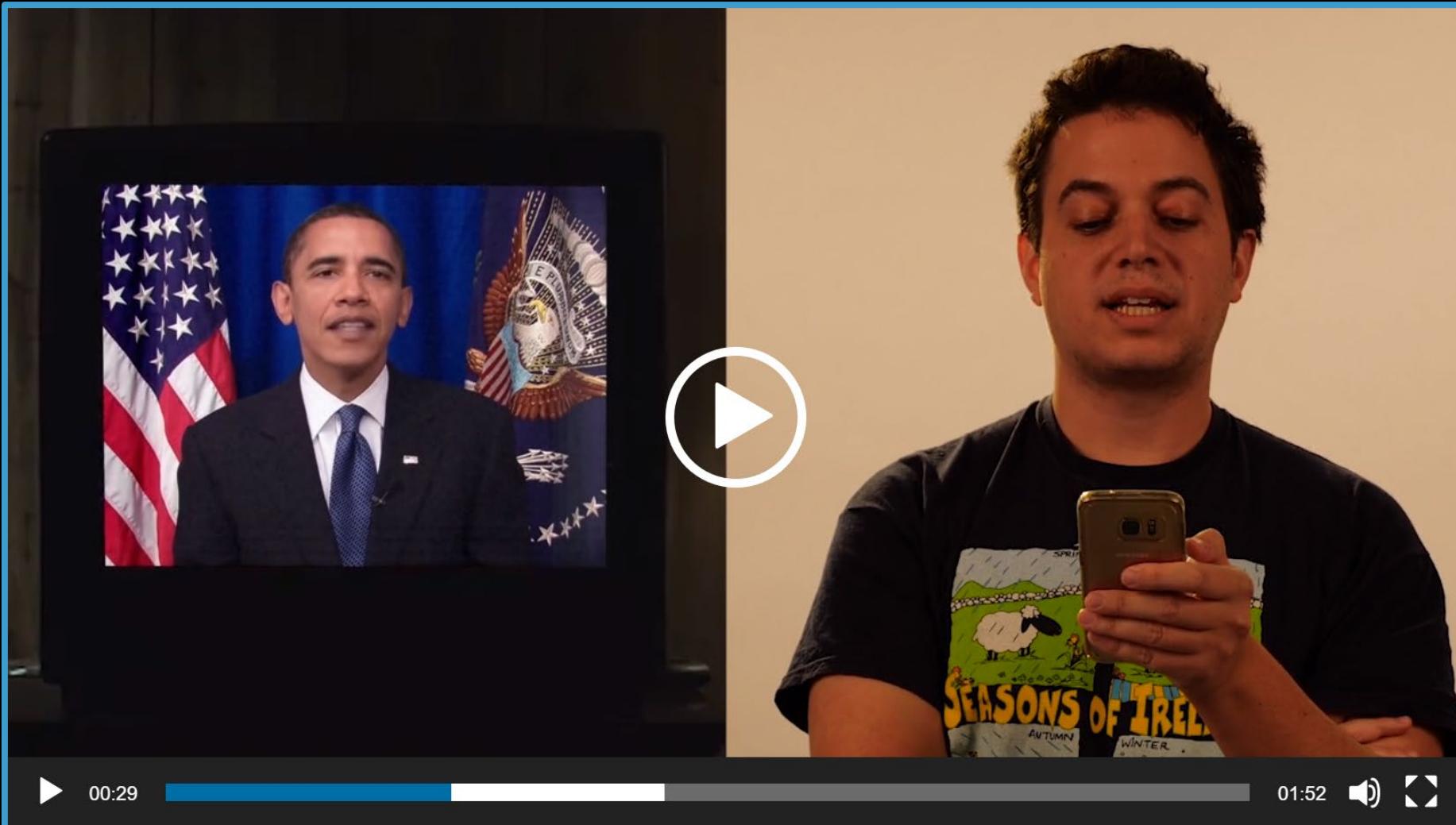


"On the Internet, nobody knows you're a dog."

- cartoon by Peter Steiner in New Yorker, 1993

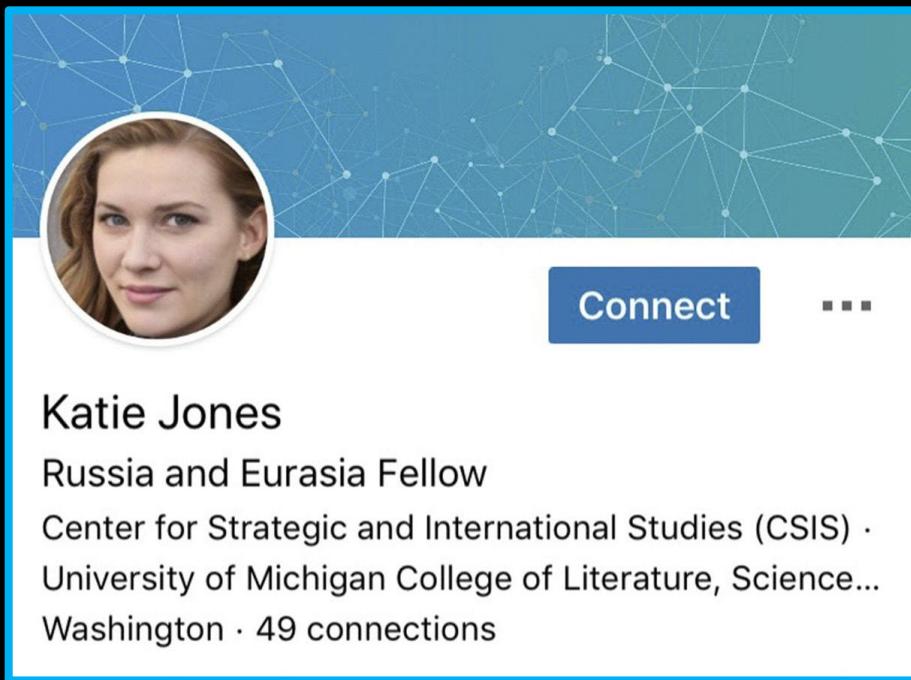


Canny AI, Imagine (2020)

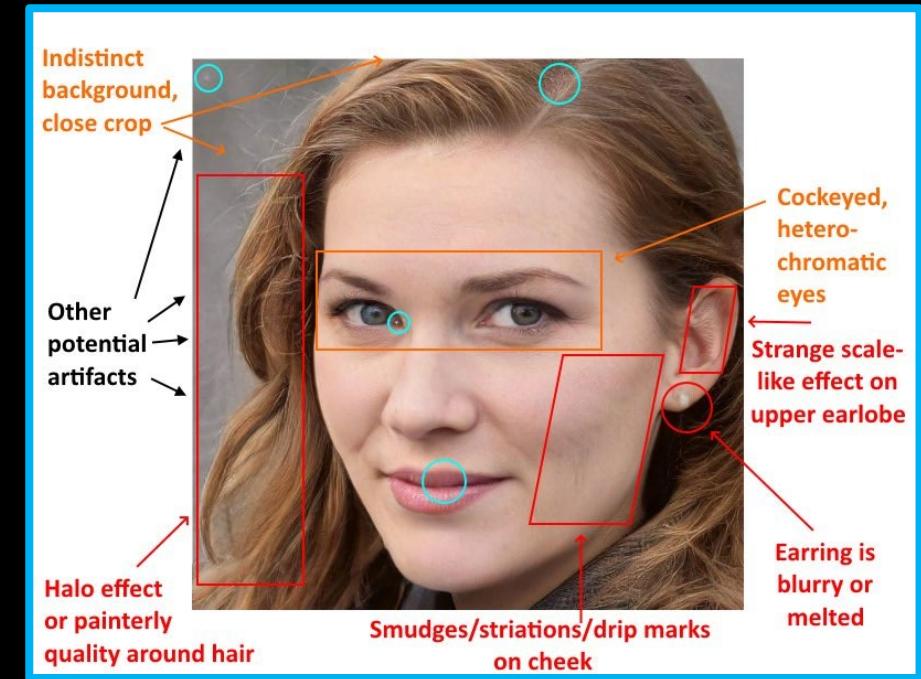


Watch: youtube.com/watch?v=KHMNPjkd5-0

Fake identity for spying (article, 2019)



A LinkedIn profile card for Katie Jones. It features a circular profile picture of a woman with blonde hair. Below the picture is her name, "Katie Jones", followed by her title, "Russia and Eurasia Fellow", and her affiliation, "Center for Strategic and International Studies (CSIS) · University of Michigan College of Literature, Science...". At the bottom, it shows "Washington · 49 connections". To the right of the profile picture are two buttons: a blue "Connect" button and a grey "...".



by Raphael Satter, journalist at Reuters, [@razhael](#)

"...The **fake profile**, given the name Katie Jones, **connected with a number of policy experts in Washington**. These included a scattering of government figures such as a senator's aide, a deputy assistant secretary of state, and Paul Winfree, an economist currently being considered for a seat on the Federal Reserve."

Overview

Deepfakes (*pre-recorded lecture*):

- **History of fake images**
- Generating fake faces **as images** and **as videos**
- **Detection**
- **Artworks**

Practical session (*during the live session*):

- **Code:** “*First Order Motion Model for Image Animation*” model

History of Fake images

“Louis Daguerre, honoured for his invention of the daguerreotype photographic process by the French Academy of Sciences and the Académie des Beaux Arts in 1839. But what about Daguerre’s contemporary Hippolyte Bayard, who had also been developing and refining his own form of photography?”



Self portrait as a drowned man, by Hippolyte Bayard (1840)

“Louis Daguerre, honoured for his invention of the daguerreotype photographic process **by the French Academy of Sciences and the Académie des Beaux Arts in 1839**. But what about Daguerre’s contemporary Hippolyte Bayard, who had also been developing and refining his own form of photography?”

“**The Government** which has been only too generous to Monsieur Daguerre, has said it can do nothing for Monsieur Bayard, and the poor wretch has drowned himself,” reads the note on the back of the photograph. “Oh the vagaries of human life....!”

[source](#)



Self portrait as a drowned man, by Hippolyte Bayard (1840)

“Louis Daguerre, honoured for his invention of the daguerreotype photographic process **by the French Academy of Sciences and the Académie des Beaux Arts in 1839**. But what about Daguerre’s contemporary Hippolyte Bayard, who had also been developing and refining his own form of photography?”

“**The Government** which has been only too generous to Monsieur Daguerre, has said it can do nothing for Monsieur Bayard, and the poor wretch has drowned himself,” reads the note on the back of the photograph. “Oh the vagaries of human life....!”

- Photograph as **proof of reality**
- Manipulation as a **statement**

[source](#)

Generating fake images

This _____ doesn't exist



This _____ doesn't exist

thispersondoesnotexist.com



thisjellyfishdoesnotexist.com

This _____ doesn't exist

thispersondoesnotexist.com



thisjellyfishdoesnotexist.com

This _____ doesn't exist

thishorsedoesnotexist.com

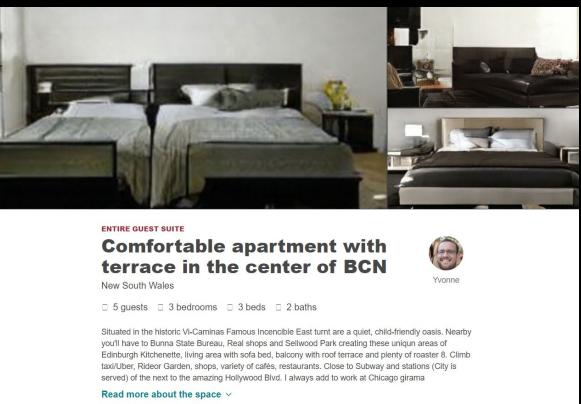


thispersondoestnotexist.com



thisjellyfishdoesnotexist.com

This _____ doesn't exist



thisrentaldoesnotexist.com

thishorsedoesnotexist.com

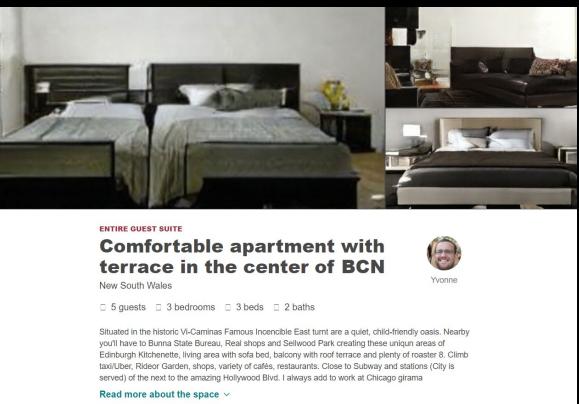


thispersondoestnotexist.com



thisjellyfishdoesnotexist.com

This _____ doesn't exist



thisrentaldoesnotexist.com

thischairdoesnotexist.com



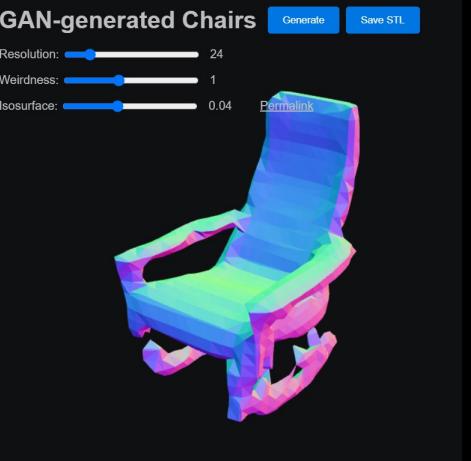
thishorsedoesnotexist.com



thispersondoestnotexist.com



thischairdoesnotexist.com



This _____ doesn't exist

thisflagdoesnotexist.com



The Flag of Lego



ENTIRE GUEST SUITE
Comfortable apartment with terrace in the center of BCN

New South Wales

5 guests 3 bedrooms 3 beds 2 baths

Situated in the historic Vi-Caminas Famous Incincible East turni are a quiet, child-friendly oasis. Nearby you'll have to Bunina State Bureau, Real shops and Selwood Park creating these unique areas of Edwin's. Rideon Garden, shops, variety of cafés, restaurants. Close to Subway and stations (City is served) of the next to the amazing Hollywood Blvd. I always add to work at Chicago, ghamas

[Read more about the space](#)

thishorsedoesnotexist.com



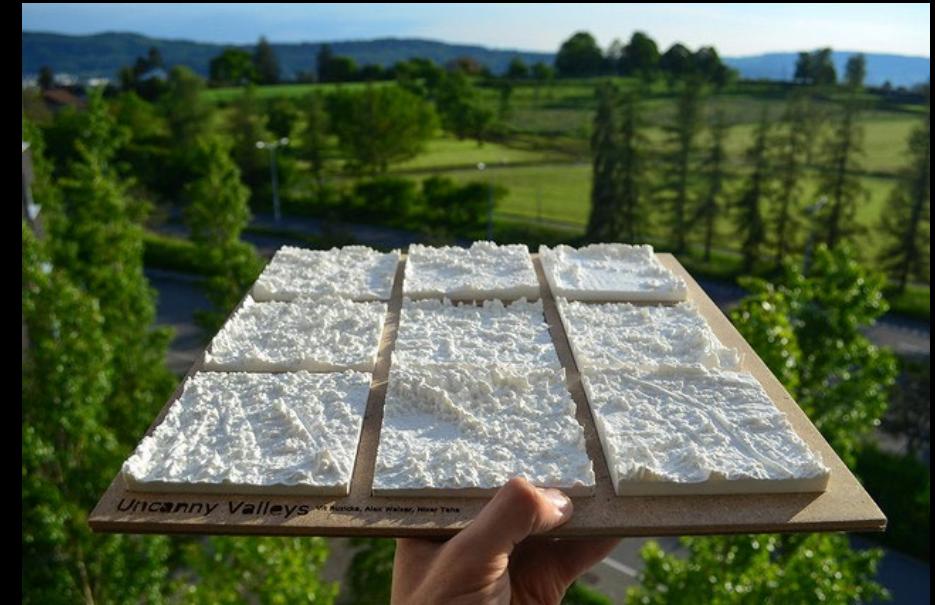
thisrentaldoesnotexist.com

This _____ doesn't exist

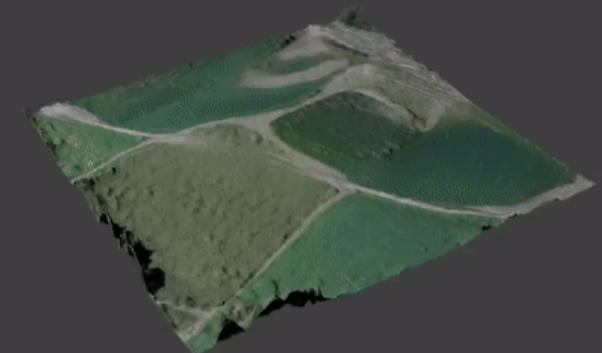
> thisxdoesnotexist.com <

A list of all similar projects

What you do with it afterwards also matters ...



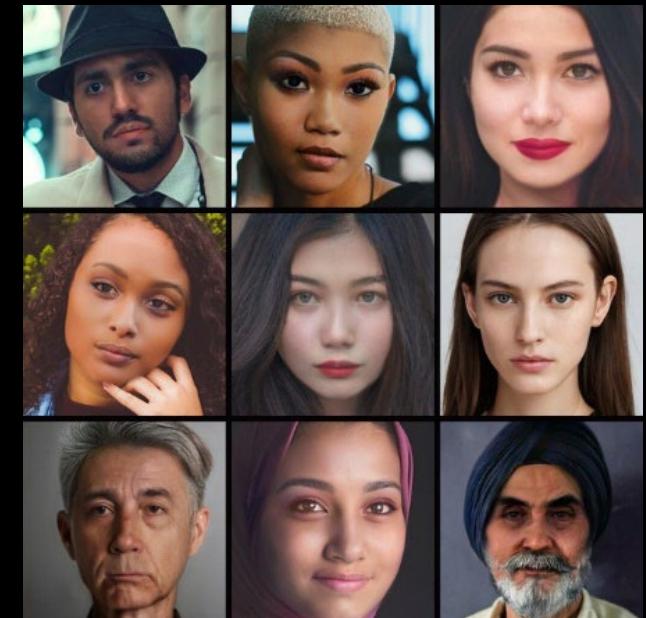
Uncanny Valleys: Generative landscape



Uncanny Valleys: Generative landscape
[online gallery](#), 2019
postcards, generated videos, 3D printed

This _____ doesn't exist

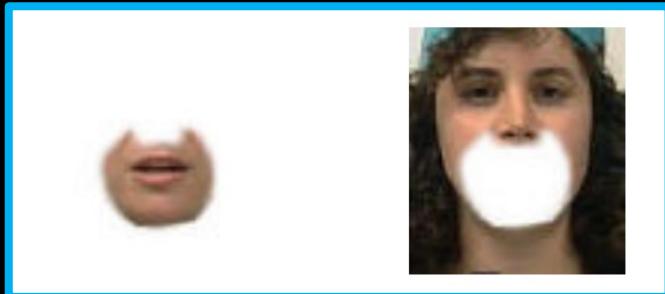
- This person doesn't exist!
- This person does exist!



Humans of AI (2020)

- Generating more than just the face
- [project](#)

Realism over years



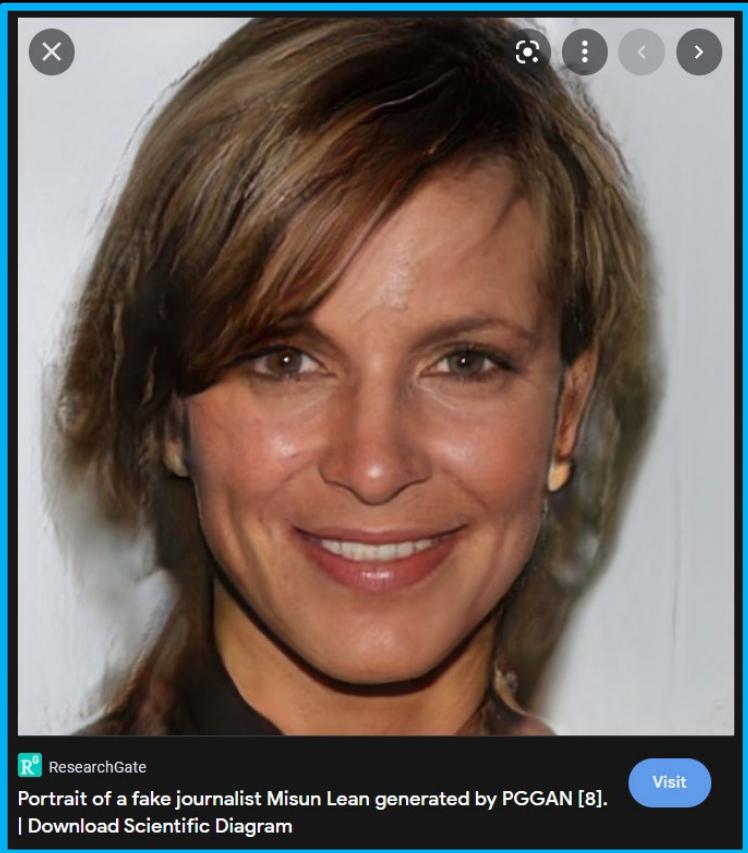
*Early work from 1997!
The [Video Rewrite](#) program
(tracking points of mouth)*



- Deep learning models getting **more and more realistic**
 - DCGAN, Progressively Growing GAN, StyleGAN v1, v2, v3, ...
 - *(In contrast: Generating the whole face)*

Fake identities

- These models have been used to **generate fake identities:**
 - Fake journalists
 - Fake attractive photos
 - Fake professional photo



Fake News, Myths of our Time
Generated journalist
part of an art project (2018),
openly described as a potential
thread (+ also generated
imagery as a reflection to the
society)



Account from 2019, covered
by Raphael Satter, journalist
at Reuters, [@razhael](#)

The websites featured articles pushing Russian talking points like “Zelensky is building a neo-Nazi dictatorship in Ukraine” and “Why Ukraine will only get worse.” As of Sunday night, the sites still featured the biographies and computer-generated faces of the columnists and linked out to their accounts on VKontakte, Russia’s Facebook competitor.



Quick thread:

I want you all to meet Vladimir Bondarenko.

He's a blogger from Kiev who really hates the Ukrainian government.

He also doesn't exist, according to Facebook.

He's an invention of a Russian troll farm targeting Ukraine. His face was made by AI.



Ben Collins  @oneunderscore_ · Feb 28

“Vladimir” has a whole backstory on the Ukraine Today website. He was an aviation engineer, until he was forced into blogging when Ukraine’s aviation infrastructure “collapsed.”

He also has weird ears, which is what happens when you make a face on [ThisPersonDoesNotExist.com](#)



Account from 2021, covered By Ben Collins, reporter for NBCNews, [@oneunderscore](#)

Control via attributes

- **Generative models which can be controlled via attributes** available in the dataset (conditional GANs)
 - One can take these attributes as a slider and change these for the generated face – for example: change age, change *gender**
 - *) *But this is mostly what has been labeled in the dataset*

Control via attributes

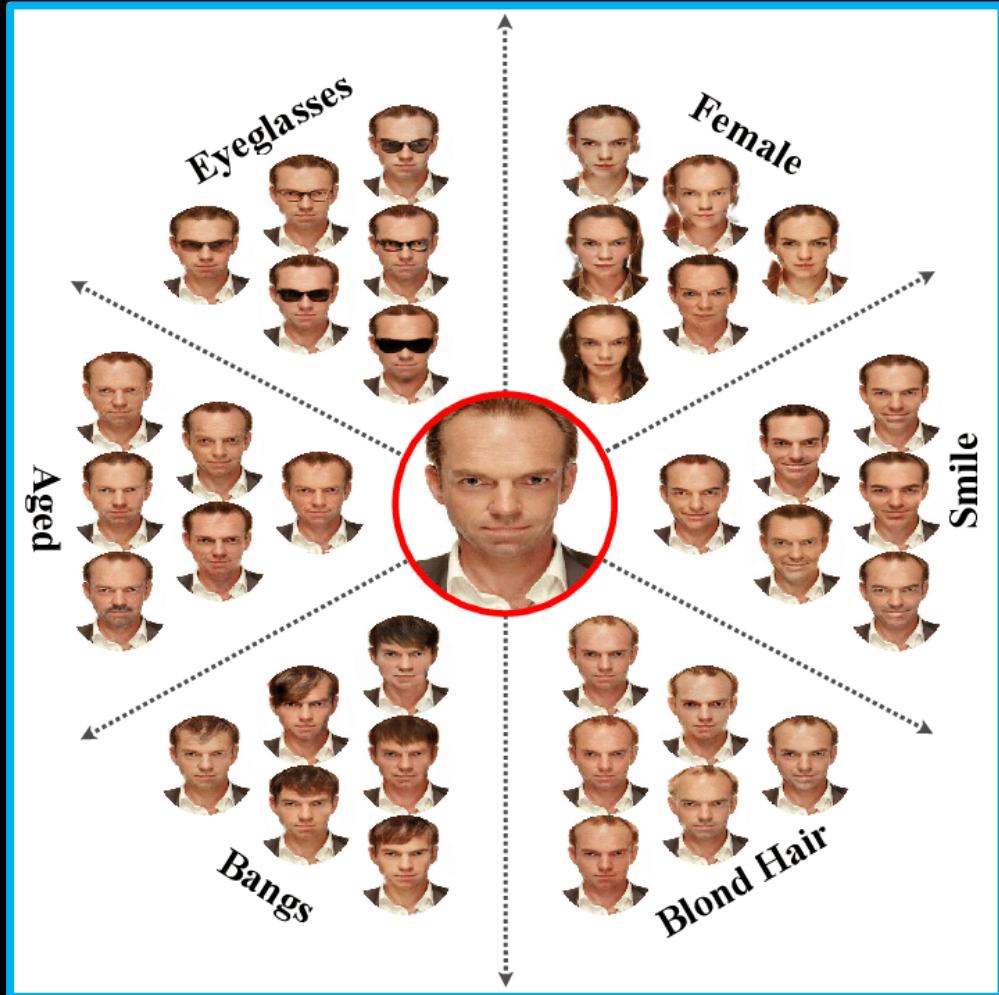
- **Generative models which can be controlled via attributes** available in the dataset (conditional GANs)
 - One can take these attributes as a slider and change these for the generated face – for example: change age, change *gender**
 - *) *But this is mostly what has been labeled in the dataset*
- With **projection**, one can embed a target (or your own) face into the model's space and edit it using these models



Control via attributes

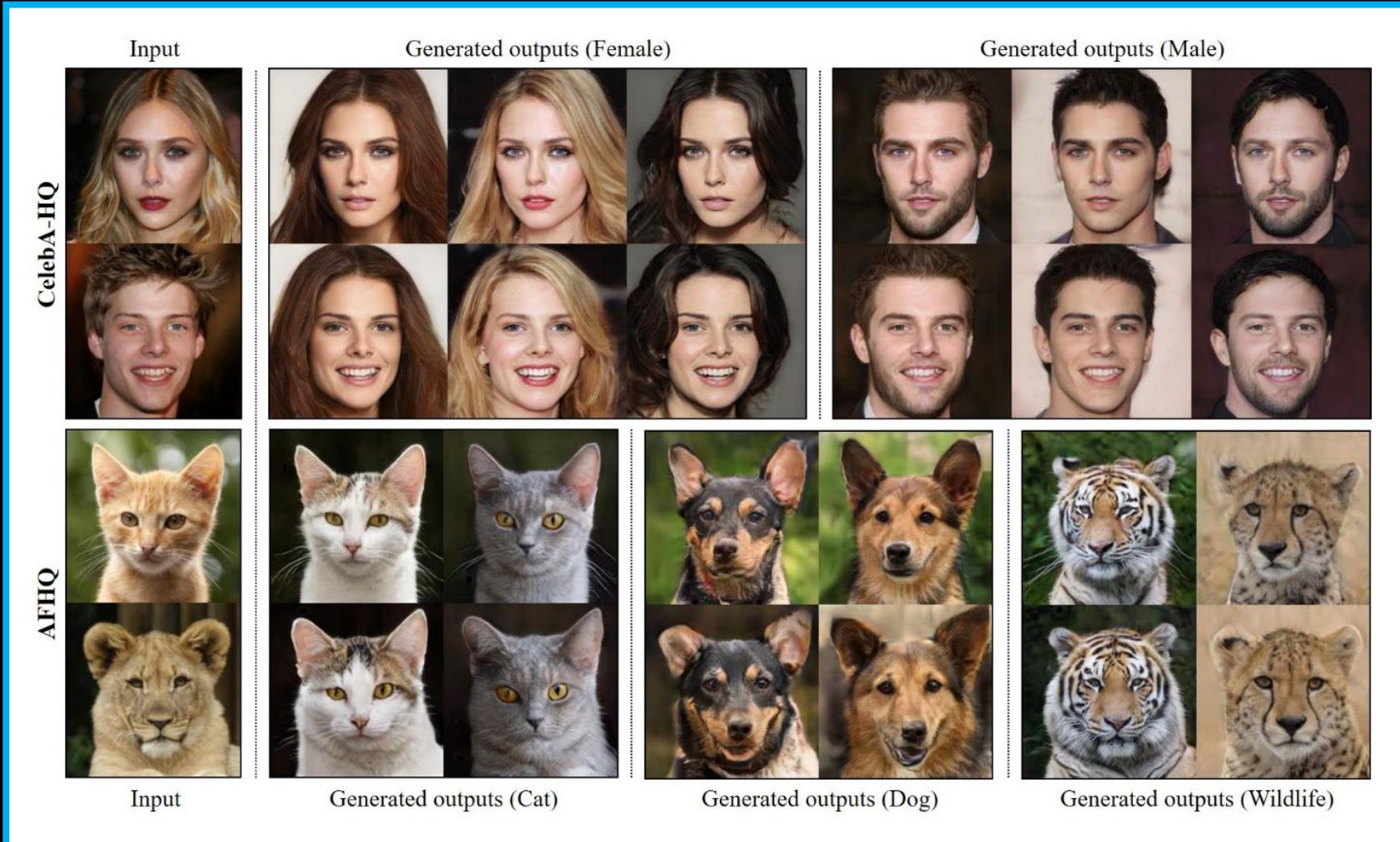
- **Generative models which can be controlled via attributes** available in the dataset (conditional GANs)
 - One can take these attributes as a slider and change these for the generated face – for example: change age, change *gender**
 - **) But this is mostly what has been labeled in the dataset*
- With **projection**, one can embed a target (or your own) face into the model's space and edit it using these models
- *Note, there are applications which can do this, but there's always a concern with sending one's likeness with these*
 - *I'd recommend trying some online Colabs instead!*

Visual attribute vectors



- **Visual attribute vectors:**
 - Eyeglasses vector
 - Smile vector
 - Blond hair vector
 - Bangs vector
 - Age vector
 - “Female” vector
- *PS: Heavily depends on what we labelled “smile”, “blond” ... or “female” ...*

StarGAN v2

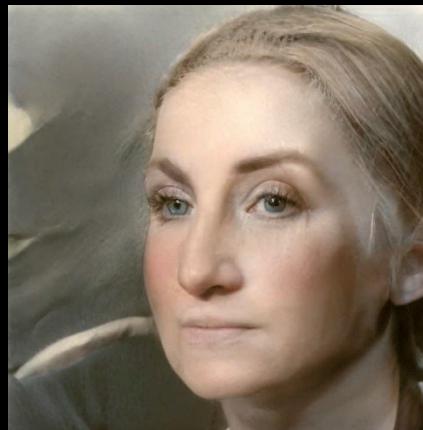


[2020] StarGAN v2: Diverse Image Synthesis for Multiple Domains, [code](#)

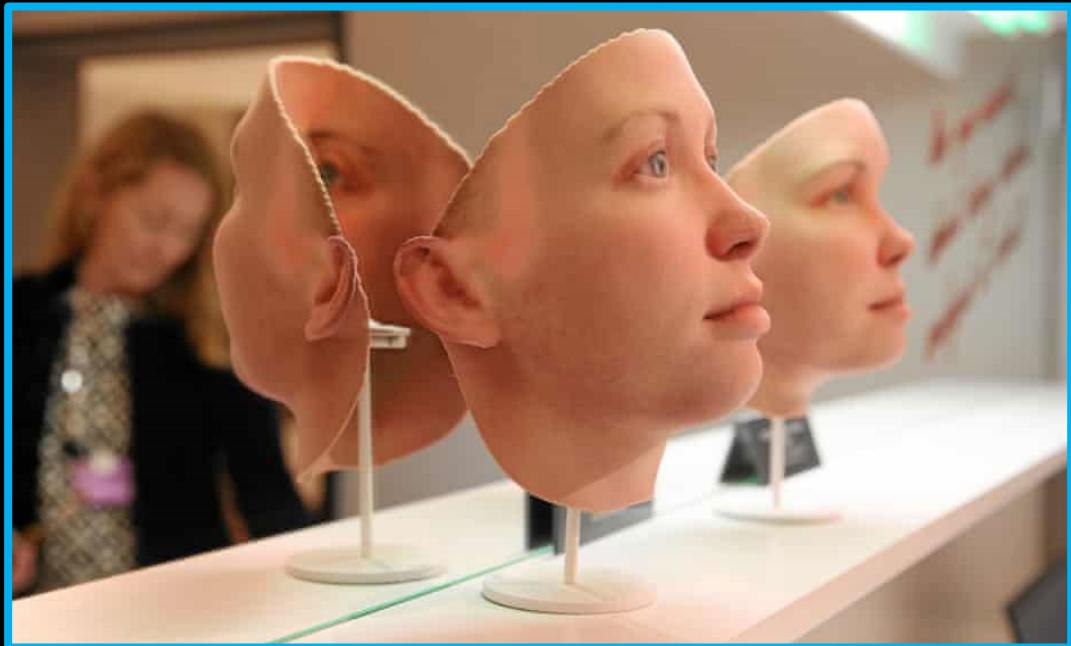
Art project: Madam president



Madam President 2020 (pre-election) reimagines all of the US Presidents as females.
- [project](#)



DNA 2 Face



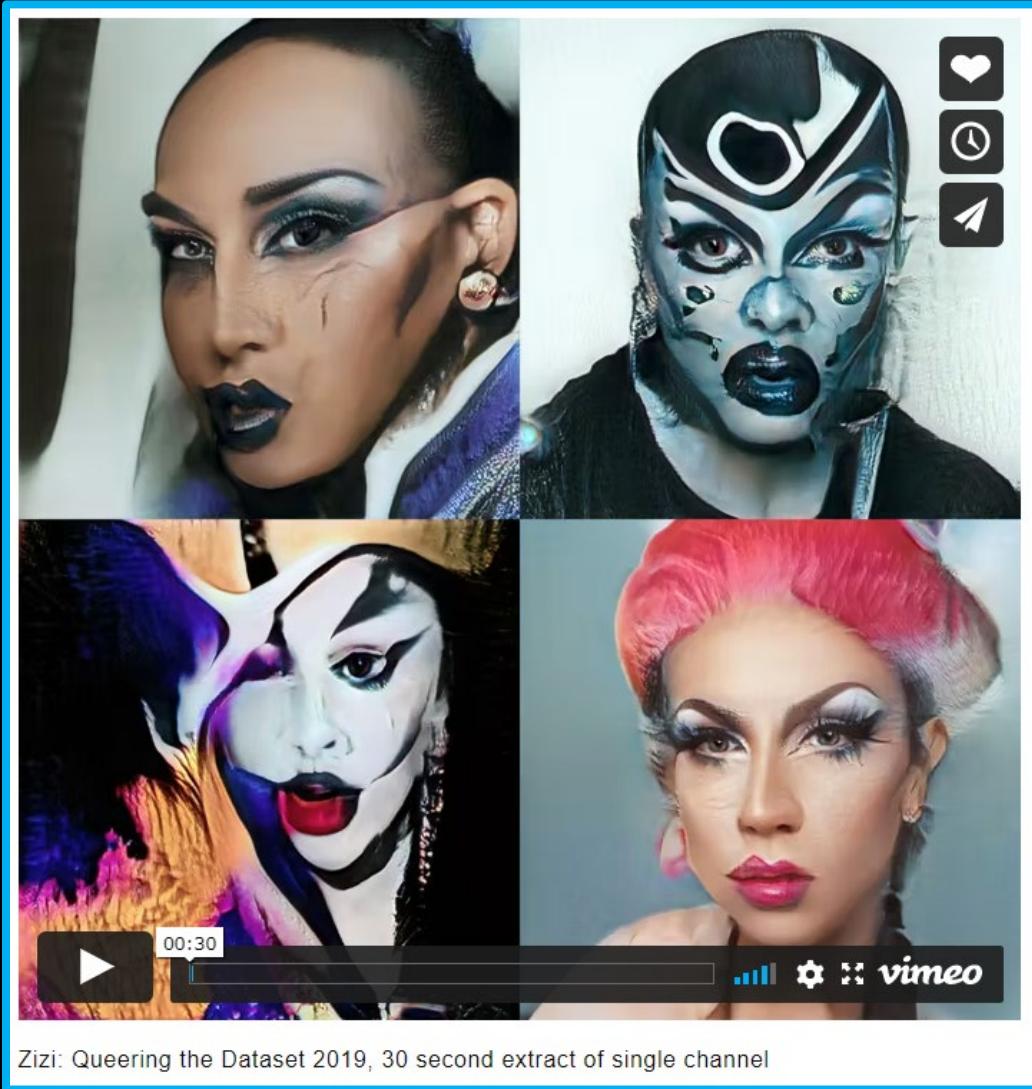
Chelsea Manning's face created by artist Heather Dewey-Hagborg

- DNA 2 Face as a forensics analysis software
- From randomly found (and also targeted) DNA samples recreated corresponding faces
- Press: [the Guardian](#), [NY Times](#)

"Feeding in those different parameters, I could generate random variations of Chelsea's face within a prescribed typology," Dewey-Hagborg told the Guardian.

The technique is the same controversial process of "forensic DNA phenotyping" that is increasingly being used by police departments to produce likenesses of suspects. The artist's work is conceived partly as a **critique of the phenotyping itself as a form of policing**.

Zizi - Queering the Dataset



Zizi: Queering the Dataset 2019, 30 second extract of single channel

- Art project playing with gender of the faces
- Perhaps these models and datasets can be used for more than propagating bias...
 - *“Made by disrupting these systems* and re-training them with the addition of 1000 images of drag and gender fluid faces found online.”*
- Project:
<https://www.jakeelwes.com/project-zizi-2019.html>

Generating fake videos

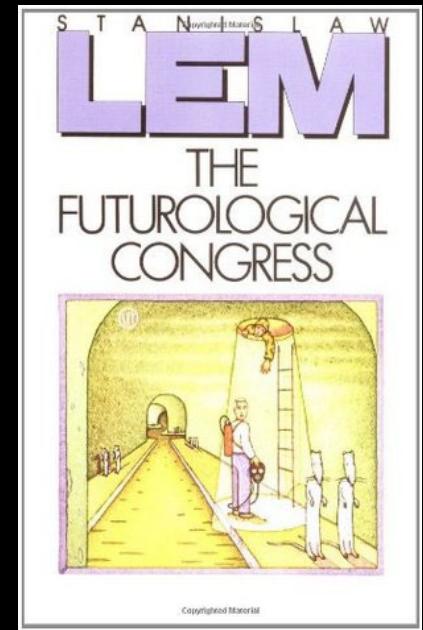
Video

- **Believability:** People are even more likely to believe a video, than a photograph
- Possible **usage in film industry**



“De-aged” examples:

Film: **The Congress** (2013)

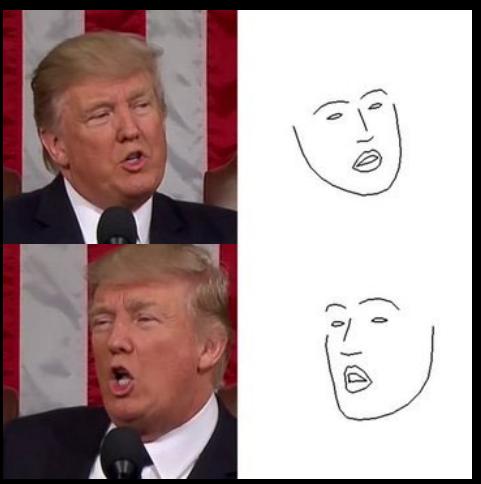




First
adapted
models

Gene Kogan's - Meat Puppets (2017), **pix2pix**

<https://twitter.com/genekogan/status/857922705412239362>



First
adapted
models

Gene Kogan's - Meat Puppets (2017), **pix2pix**

<https://twitter.com/genekogan/status/857922705412239362>

“**CycleGAN** Face-off (2017)” ([Project](#))

- CycleGAN allows unpaired faces



Advanced models



First Order Motion Model for Image Animation (2019)

<https://github.com/AliaksandrSiarohin/first-order-model>

Needs:

- Driving video for **motion**
- Single image for the **face**

*Follow-up research:
Motion Representations for
Articulated Animation (2021)*

Alongside an exhibition

"Dalí once said: *"Si muero, no muero por todo,"* or *"If I die, I won't completely die."* Thirty years after his death, his words take on a new meaning at The Dalí Museum in St. Petersburg, Florida. "Dalí Lives" uses artificial intelligence to let visitors experience his bigger-than-life personality in an up close and personal way."

Dalí Lives – Art Meets Artificial Intelligence. (2019)

Likely process:

- *Actor performance*
- *With AI swapped faces*

Deep Nostalgia

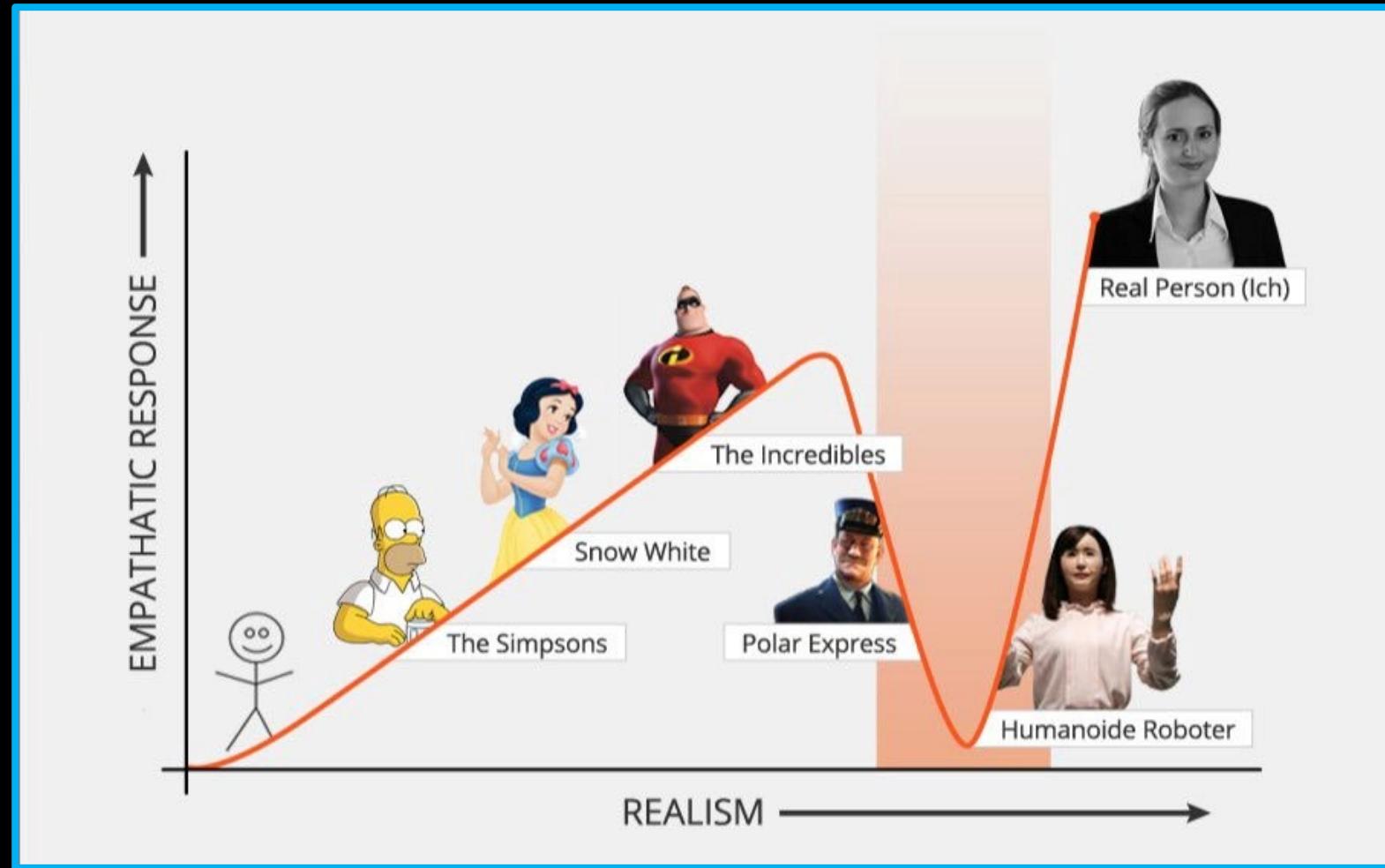


Company offering:

- Animating your relatives from old photographs
- Website: myheritage.com/deep-nostalgia

Likely process:

- *Actor performance*
- *With AI swapped faces*



Detection

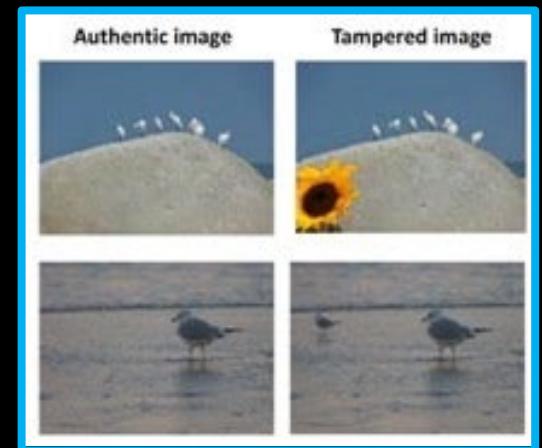
Intuition

- Generated images tend to have **specific noise**, almost like **tell-tale glitches and patterns**
- After working with GANs for a while you will also begin to see these
 - Some linked to particular model versions
 - StyleGAN v1 had strange **blobs** for example:
 - Most of GANs have **ripples** due to the Convolutions in the models

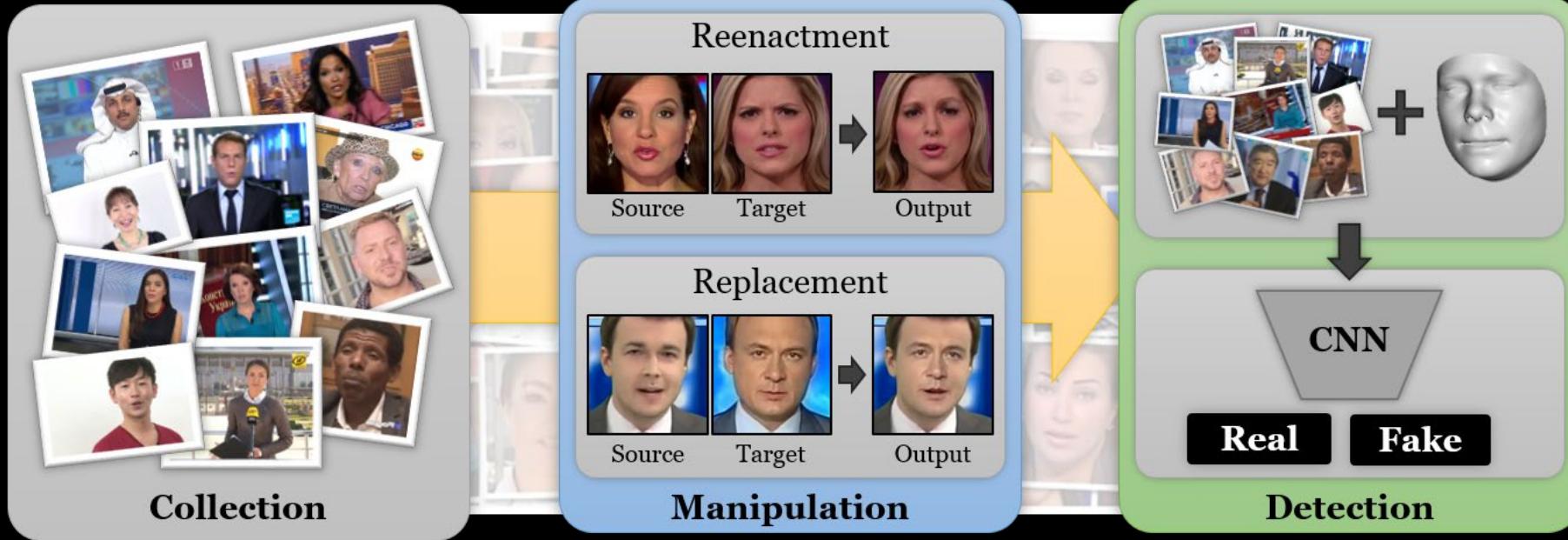


Detection as a task

- Could this detection be done automatically?
- Train a **ML model to detect ML generated images?**
 - Theoretically it's a simple classification task
 - Theoretically even when training a GAN, we have a Discriminator determining real/fake labels
- Detecting wide array of **manipulations**

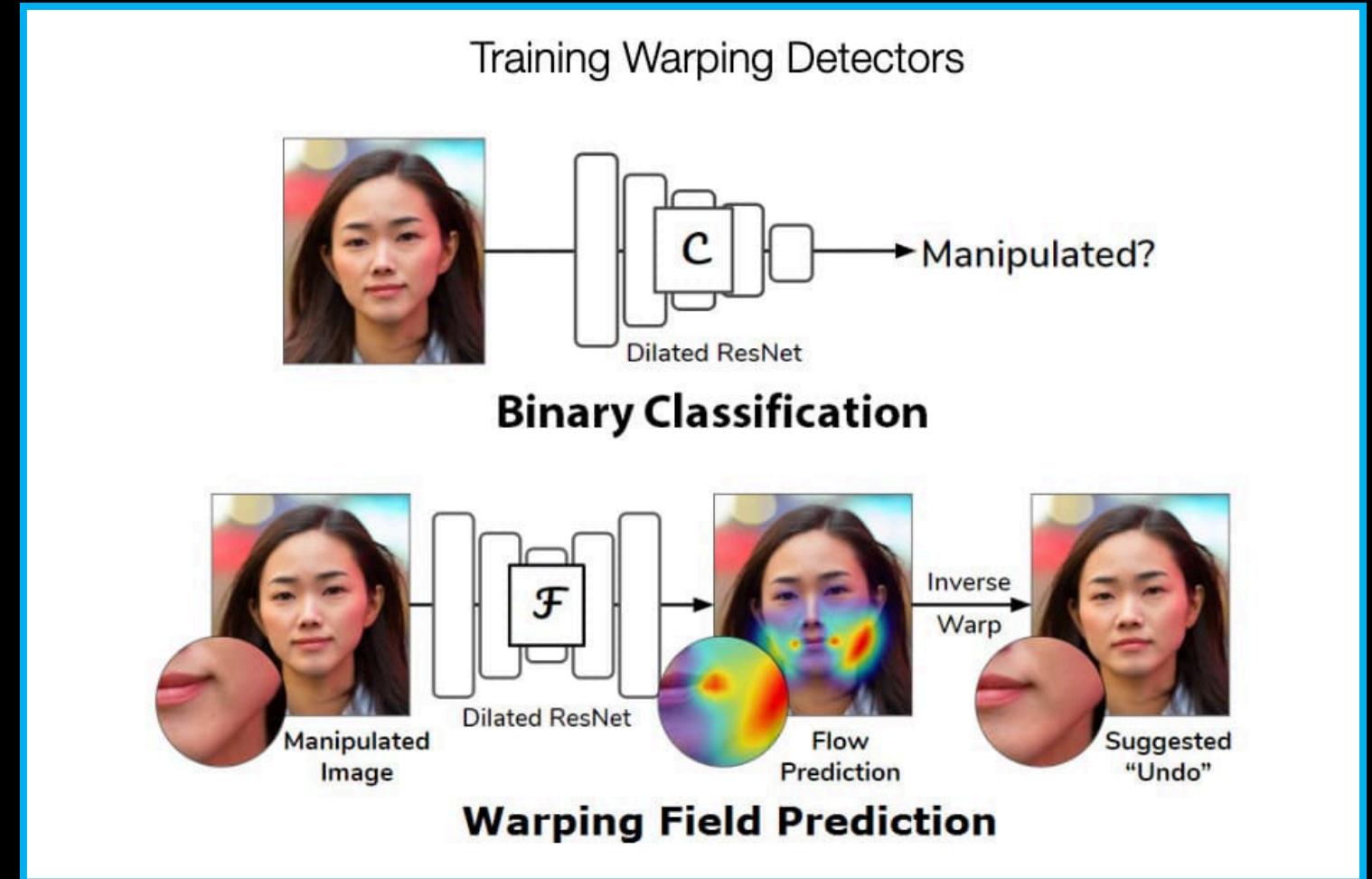


Research in the area



- There are datasets and active research in this area – because there's a real danger in not detecting fake identities
 - Dataset: github.com/ondyari/FaceForensics/tree/master/dataset

Software



- Proposed as a tool for Adobe Photoshop
 - Adobe Research [article](#) (2019), on [fxguide](#)
 - **Research** [here](#) and [here](#)

Examples

Generated identities



(we already talked about generated identities which were made
to **give more credibility to social network accounts**)

Generated identities

- **Politicians**, a usual target for **mockery** with these models, but also a **dangerous** area
 - Can a video of a politician's statement be trusted?
 - How can it be verified in the media that someone's footage is real?
 - *{Are we living in post-real world?}*

The dark side of deepfakes

- Sadly, pornography, wide spread of this at the start. Very obviously very dangerous – regardless of the fidelity of the models, it can cause large amounts of harm!
 - These videos got later banned ...

The dark side of deepfakes

- Same technology, but **safe for work**

examples:

- Actor replacement between movies
(enough data online)
- Slightly less malign examples of these methods...
- **Use this as a lesson:** Notice what works and notice what doesn't on these videos



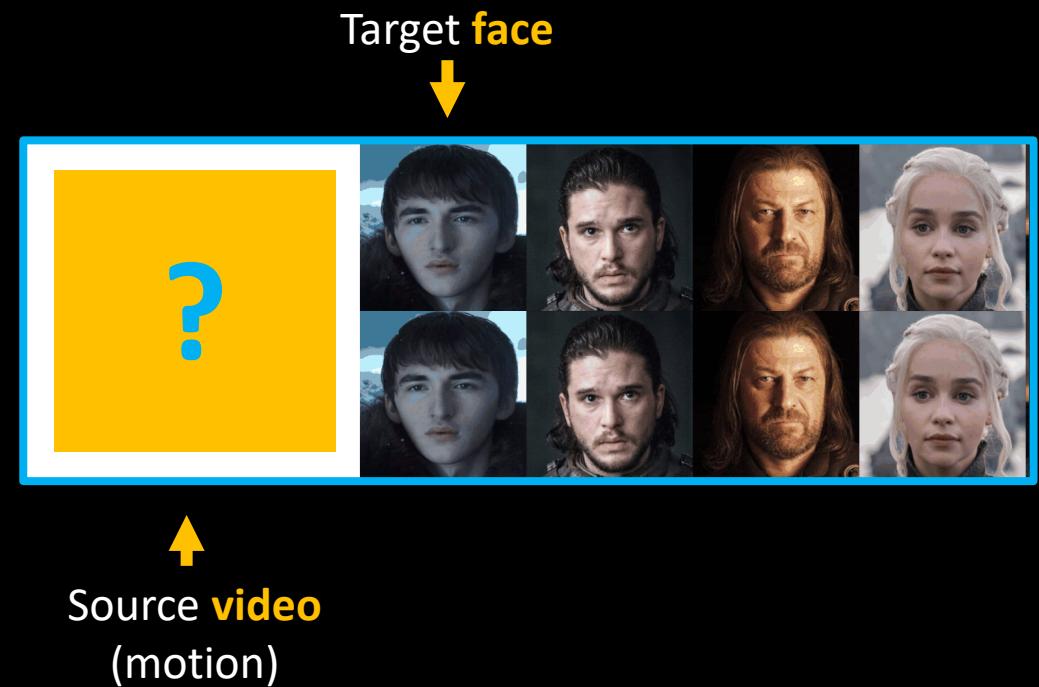
Links: [video collage](#), [r/deepcage/ community](#)

Summary from the lecture

- **Deepfakes:**
 - Methods to **generate fake images (GANs)** and **fake videos**
 - Methods to **detect manipulated data (classification)**
 - **Examples in news (attempts of misinformation), artworks (both in revealing these problems and manipulation of face datasets)**

Preparation for the practical

- Prepare to record your own **source video**
- Think about possible **target face** image
- Prepare to have your own cameras on at least at the start of the practicals



<https://www.artbreeder.com/i?k=5dc654373c0b65c21de0>



AI for the Media

Week 9, DeepFakes



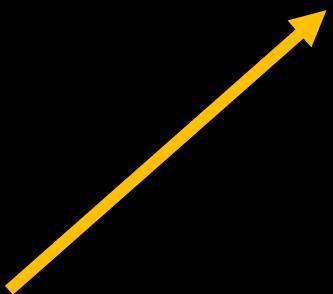
Practical: Deep Nostalgia recreation

First Order Motion Model for Image Animation

Target **face**



Source **video** (motion)



Possible use case, animating
game avatars



First Order Motion Model for Image Animation

Needs:

- Get a **source data** of movements / expressions
 - source video – your own made!
- And a **single photo of the target face**
 - Hopefully not very harmful! (low fidelity) **but please keep this in mind when working on these projects**
- Can even be playful, ex: animating Mona Lisa



Possible suggestion, getting a generated face from **artbreeder**:

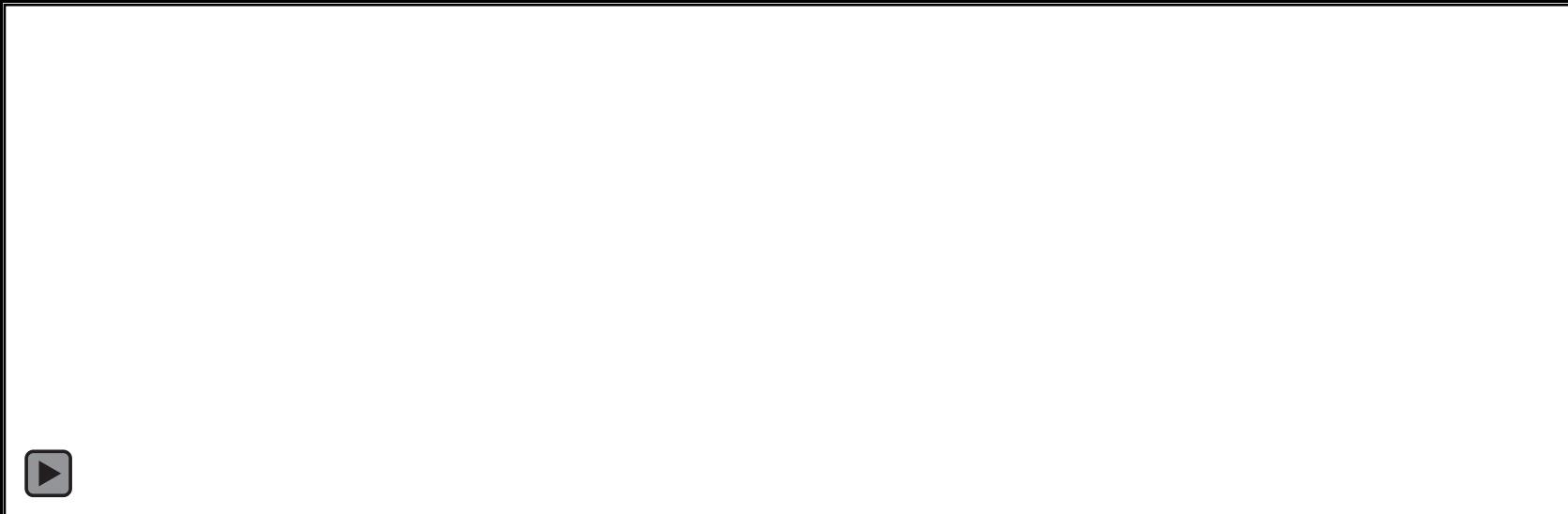


Example: artbreeder.com/i?k=65aeb101d7c882d81225

Target **face**



Result video



Source **video** (motion)



<https://www.artbreeder.com/i?k=cc4903632bd13c1cc7fc7178408c>



<https://www.artbreeder.com/i?k=920ce5b3373a70bff383d60ed7a5>



<https://www.artbreeder.com/i?k=6201da4167cd151cfb8f9b22c5f9>



<https://www.artbreeder.com/i?k=922c26efd402d232a116956fc8c4>



<https://www.artbreeder.com/i?k=f0ea6fcb50bb0282ae37b895f8f8>



<https://www.artbreeder.com/i?k=5dc654373c0b65c21de0>



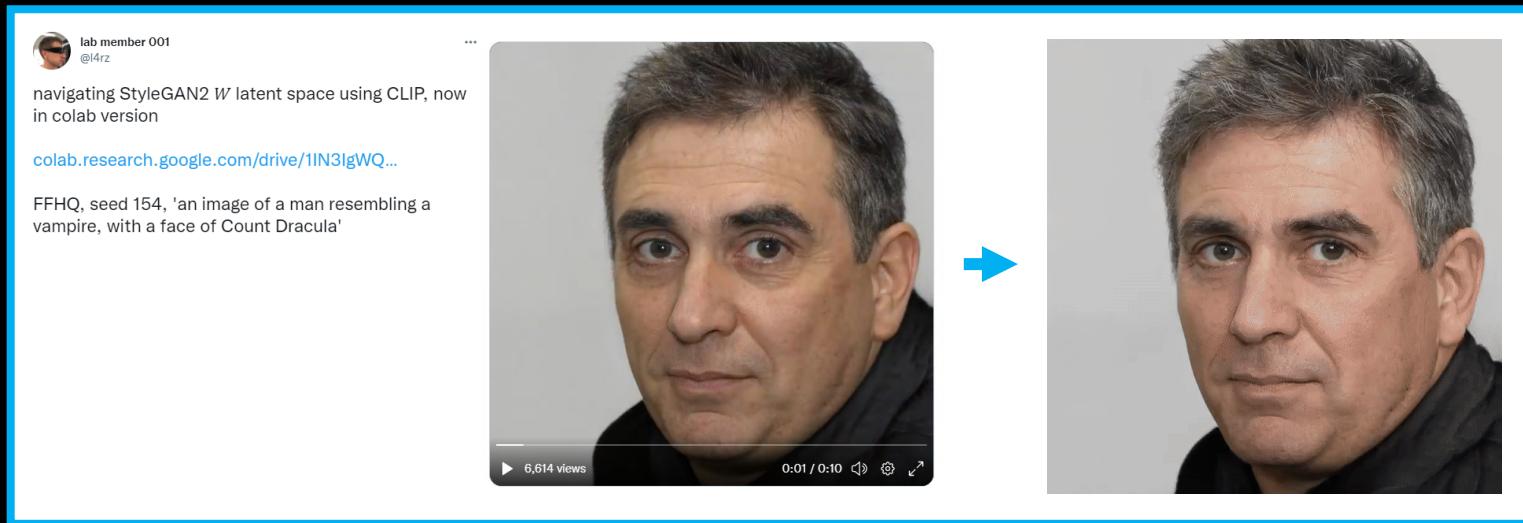
<https://www.artbreeder.com/i?k=9564cfb30bf1d5a6341f0ea6a2cb>



<https://www.artbreeder.com/i?k=6ec5463f6e28941750a2>

Methods chaining

Possible suggestion – use some of the previously introduced methods of **controlling generative models** (for example **txt2image guidance**):



Target **face**
↓



Colab [available on our github!](#)

The end