Quiz 9 (on Canvas)

Ends at 1:08pm

CS5670: Computer Vision

Computer Vision, Society, and Ethics

Announcements

- Project 5 (Neural Radiance Fields) due Weds, May 3 by 8pm
- In class final on May 9
  - Open book, open note
- Course evaluations are open starting Monday, May 1
  - We would love your feedback!
  - Small amount of extra credit for filling out
    - What you write is still anonymous, instructors only see whether students filled it out
  - Link coming soon

Additional Resources

- FATE (Fairness Accountability Transparency and Ethics) in Computer Vision Tutorial
  - Timnit Gebru and Emily Denton
  - https://sites.google.com/view/fatecv-tutorial/schedule

- https://exposing.ai/
  - Adam Harvey and Jules LaPlace

Advances in computer vision

- Sometimes we think of technological development as a uniform positive
- But computer vision exists in a societal context, and can have both good and bad consequences – need to be mindful of both
- Example: as computer vision gets better, our privacy gets worse (e.g., through improved face recognition)

Today

- Examples of bias in computer vision and beyond
- Datasets and unintended consequences
- DeepFakes and image synthesis methods

Questions

- Should I be working on this problem at all?
- Does a given vision task even make sense?
- What are the implications if it doesn’t work well?
- What are the implications if it does work well?
- What are the implications if it works well for some people, but not others?
- Who benefits and who is harmed?
- (About datasets) How was it collected? Is it representative?
- (For any technology) Who is it designed for?

More questions

- Does the application align with your values?
- Does the task specification / evaluation metric reflect the things you care about?
- For recognition:
  - Does the collected training / test set match your true distribution?
- Are the algorithm's errors biased?
- Are you being honest in public descriptions of your results?
- Is the accuracy/correctness sufficient for public release?

Slide credit: Bharath Hariharan

Bias in computer vision and beyond

- What follows are a number of examples of bias from the last 100 years

Shirley cards

Example Kodak Shirley Card,
1950s and beyond

Kodak’s Multiracial Shirley Card,
North America. 1995.

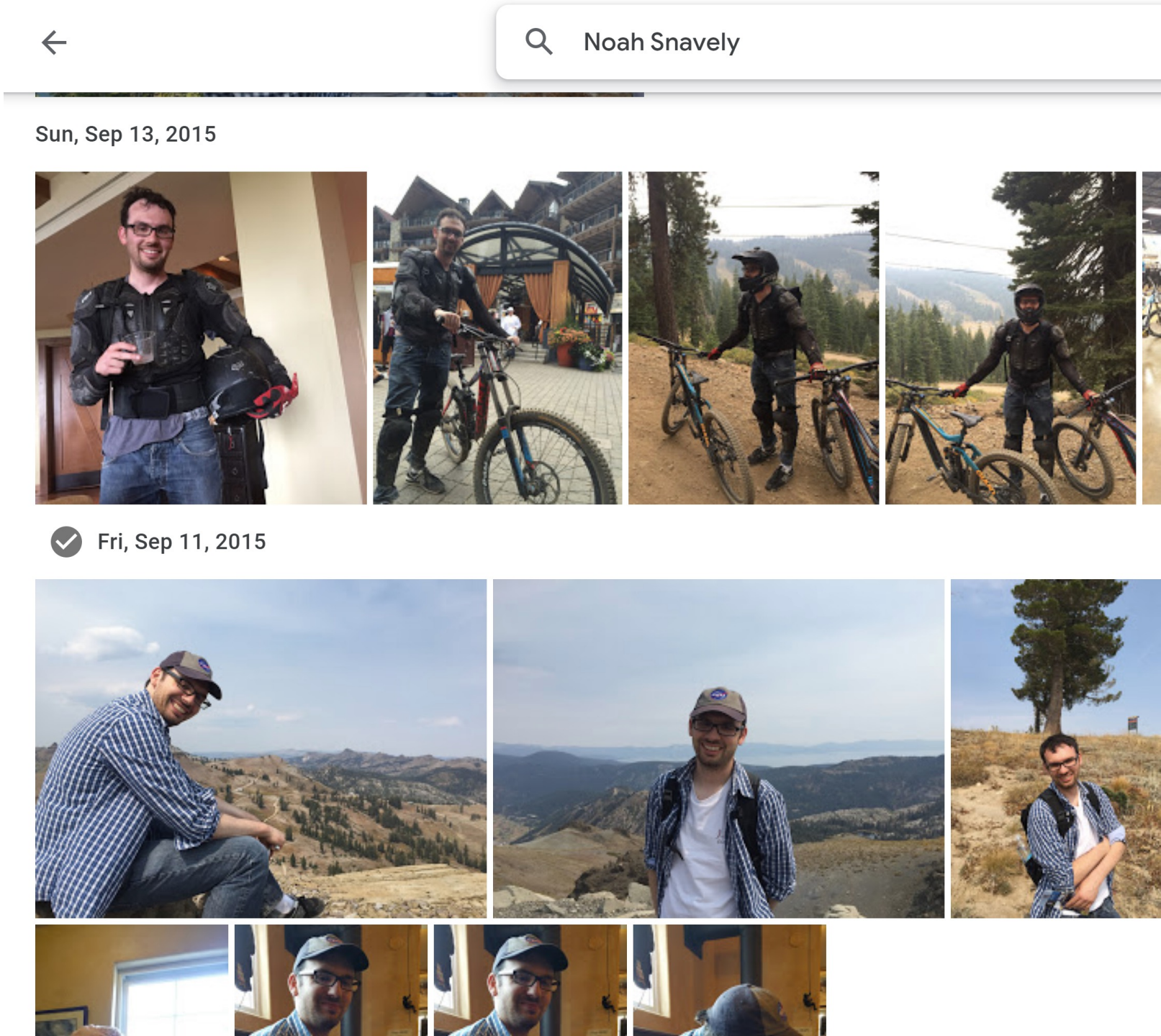
How Kodak's Shirley Cards Set Photography's Skin-Tone Standard
https://www.npr.org/2014/11/13/363517842/for-decades-kodak-s-shirley-cards-set-photography-s-skin-tone-standard

The Racial Bias Built Into Photography
https://www.nytimes.com/2019/04/25/lens/sarah-lewis-racial-bias-photography.html

Face recognition

- Probably the most controversial vision technology
- Three different versions:
  - Face verification: “Is this person Noah Snavely?” (e.g., Apple’s Face Unlock)
  - Face clustering: “Who are all the people in this photo collection”? (e.g., Google Photos search)
  - Face recognition: “Who is this person”? (e.g., identify a person from surveillance footage of a crime scene)
- Applications can suffer from bias (working well for some populations but not others) and misuse

Google Photos automatic face clustering and recognition



Many Facial-Recognition Systems Are Biased, Says U.S. Study

Algorithms falsely identified African-American and Asian faces 10 to 100 times more than Caucasian faces, researchers for the National Institute of Standards and Technology found.

Morning at Grand Central Terminal. Technology for facial recognition is frequently biased, a new study confirmed. Timothy A. Clary/Agence France-Presse — Getty Images

NYT, 12/20/19
https://www.nytimes.com/2019/12/19/technology/facial-recognition-bias.html

Account

Wrongfully Accused by an Algorithm

In what may be the first known case of its kind, a faulty facial recognition match led to a Michigan man's arrest for a crime he did not commit.

New York Times, June 24, 2020
https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html

The New York Times

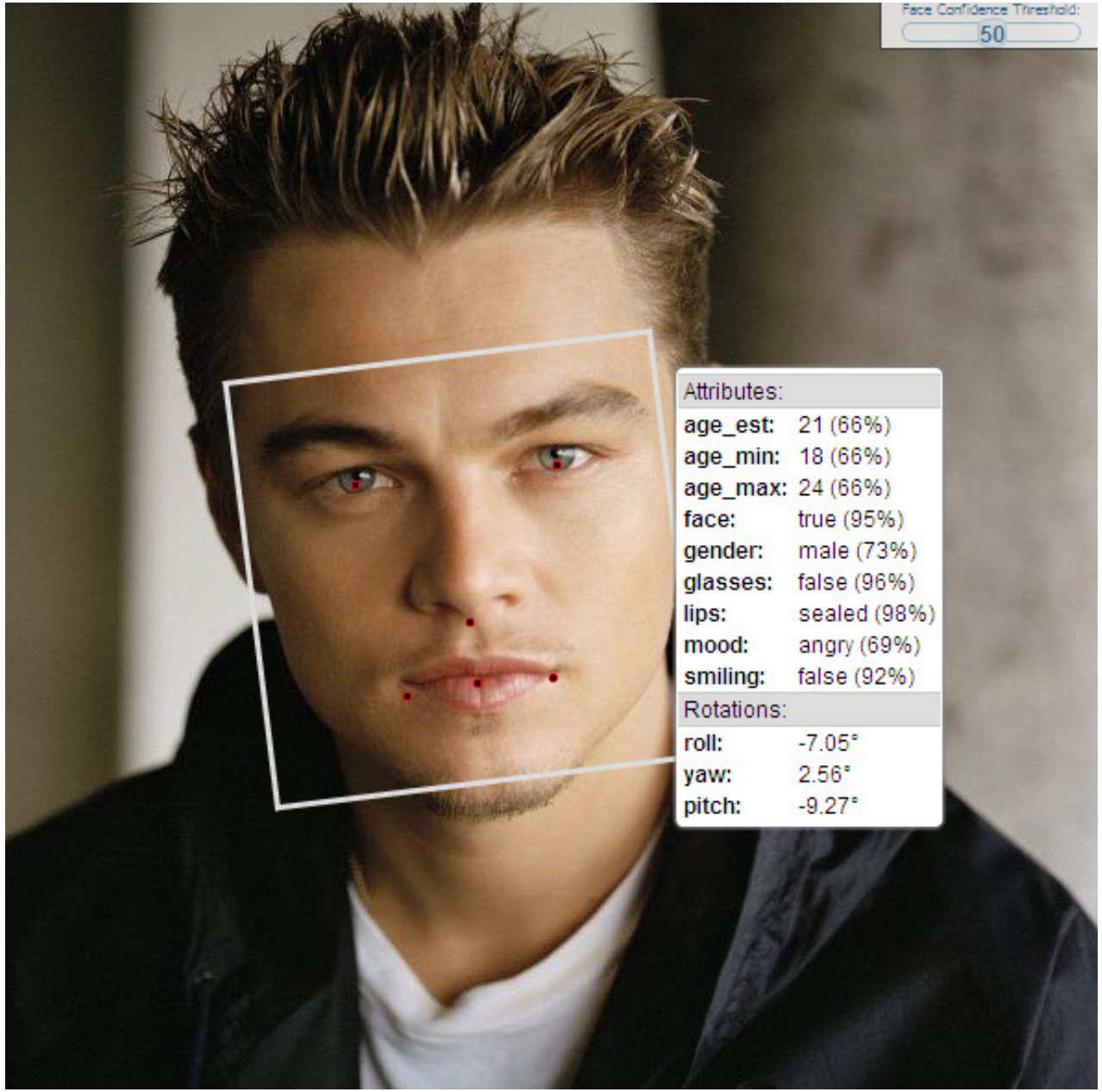
The Secretive Company That Might End Privacy as We Know It

A little-known start-up helps law enforcement match photos of unknown people to their online images — and "might lead to a dystopian future or something," a backer says.

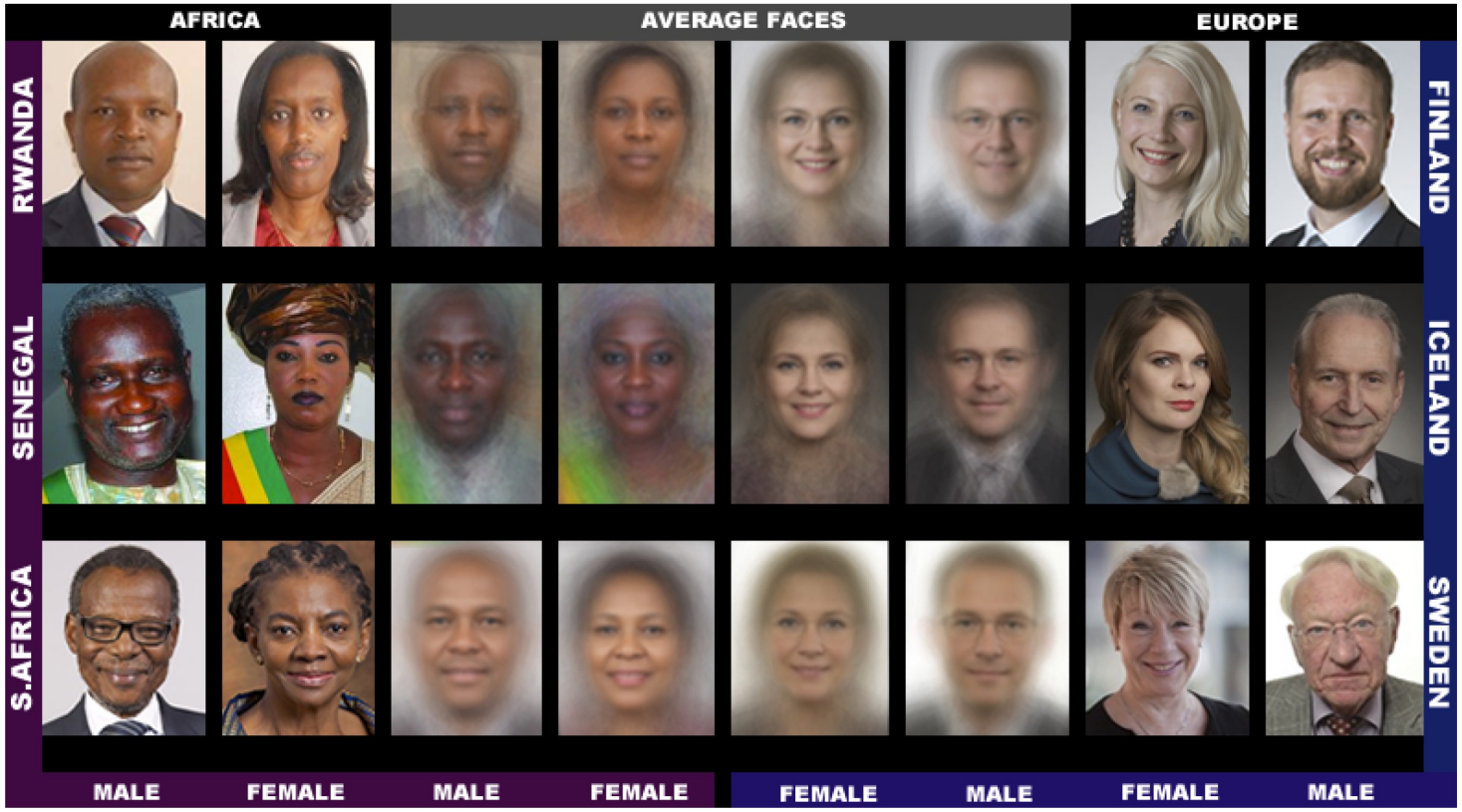
https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html

Face analysis

- Gender classification
- Age regression
- Expression classification
- Ethnicity classification



Gender Shades – Evaluation of bias in Gender Classification



Images from the Pilot Parliaments Benchmark

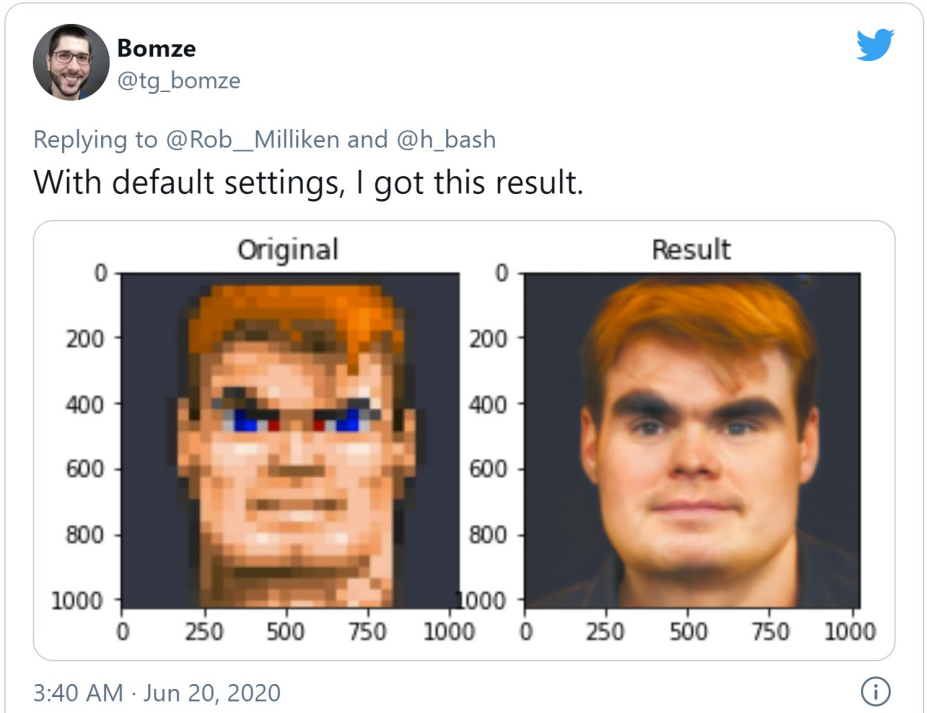
Joy Buolamwini and Timnit Gebru. **Gender shades: Intersectional accuracy disparities in commercial gender classification.** Conference on Fairness, Accountability and Transparency. 2018.

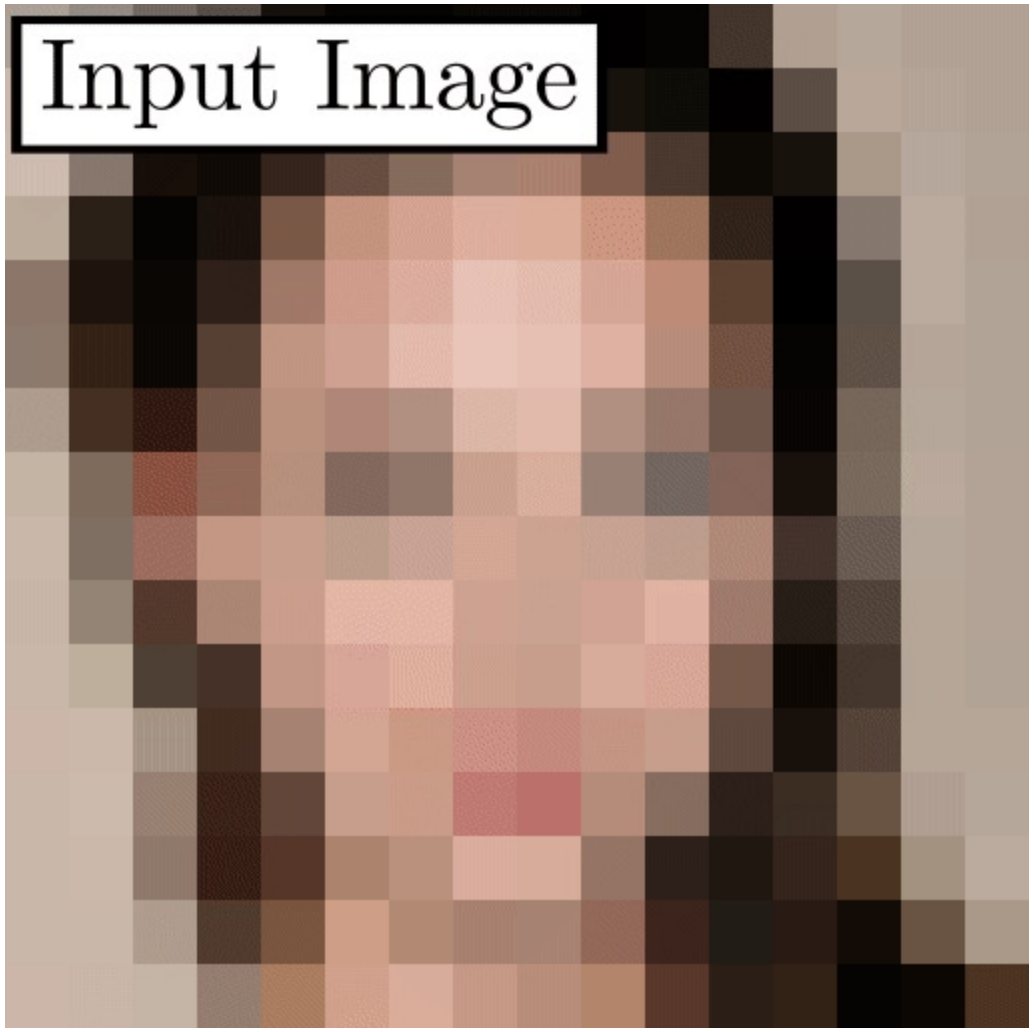
Classifier	Metric	All	F	M	Darker	Lighter	DF	DM	LF	LM
MSFT	PPV(%)	93.7	89.3	97.4	87.1	99.3	79.2	94.0	98.3	**100**
	Error Rate(%)	6.3	10.7	2.6	12.9	0.7	**20.8**	6.0	1.7	0.0
	TPR (%)	93.7	96.5	91.7	87.1	99.3	92.1	83.7	**100**	98.7
	FPR (%)	6.3	8.3	3.5	12.9	0.7	**16.3**	7.9	1.3	0.0
Face++	PPV(%)	90.0	78.7	99.3	83.5	95.3	65.5	**99.3**	94.0	99.2
	Error Rate(%)	10.0	21.3	0.7	16.5	4.7	**34.5**	0.7	6.0	0.8
	TPR (%)	90.0	98.9	85.1	83.5	95.3	98.8	76.6	**98.9**	92.9
	FPR (%)	10.0	14.9	1.1	16.5	4.7	**23.4**	1.2	7.1	1.1
IBM	PPV(%)	87.9	79.7	94.4	77.6	96.8	65.3	88.0	92.9	**99.7**
	Error Rate(%)	12.1	20.3	5.6	22.4	3.2	**34.7**	12.0	7.1	0.3
	TPR (%)	87.9	92.1	85.2	77.6	96.8	82.3	74.8	**99.6**	94.8
	FPR (%)	12.1	14.8	7.9	22.4	3.2	**25.2**	17.7	5.20	0.4

Joy Buolamwini and Timnit Gebru. **Gender shades: Intersectional accuracy disparities in commercial gender classification.** Conference on Fairness, Accountability and Transparency. 2018.

Case study – upsampling faces

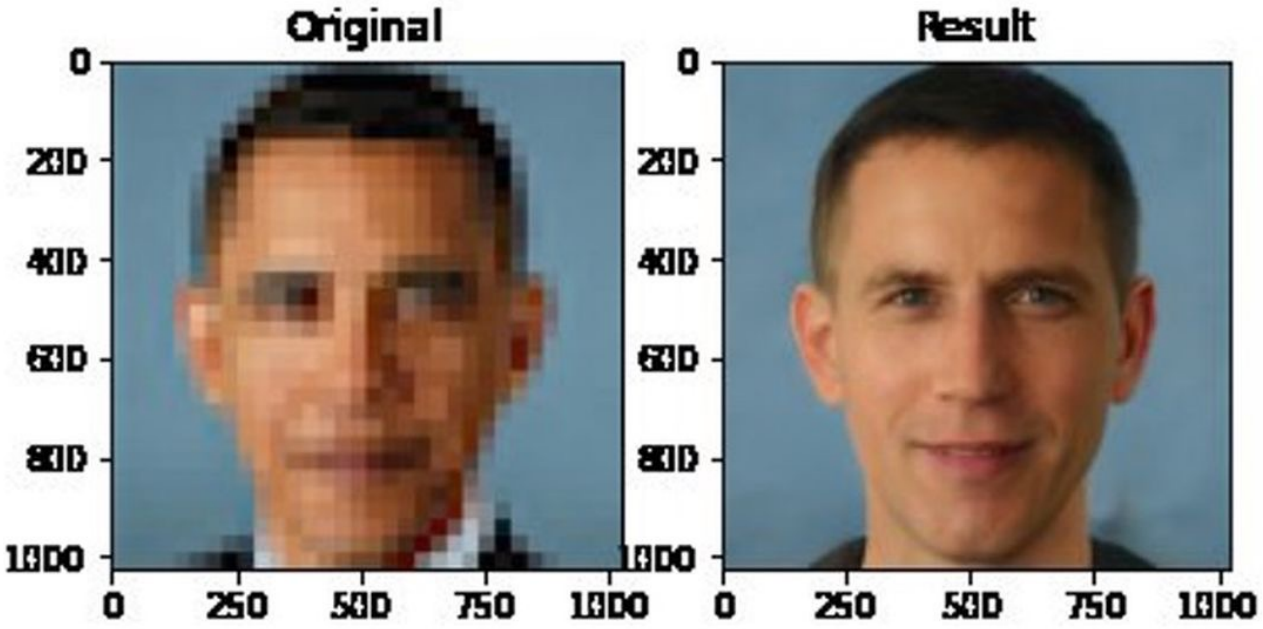
PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models
Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, and Cynthia Rudin
https://arxiv.org/abs/2003.03808





https://github.com/tg-bomze/Face-Depixelizer

Case study – upsampling faces



https://www.theverge.com/21298762/face-depixelizer-ai-machine-learning-tool-pulse-stylegan-obama-bias

Case study – upsampling faces

"We have noticed a lot of concern that PULSE will be used to identify individuals whose faces have been blurred out. We want to emphasize that this is impossible - PULSE makes imaginary faces of people who do not exist, which should not be confused for real people. It will not help identify or reconstruct the original image.

We also want to address concerns of bias in PULSE. We have now included a new section in the paper and an accompanying model card directly addressing this bias."

https://github.com/tg-bomze/Face-Depixelizer, accessed May 4, 2021

Case study – classifying sexual orientation

0.0B | Public | 16 | ...

Deep neural networks are more accurate than humans at detecting sexual orientation from facial images.

Contributors: Yilun Wang, Michal Kosinski

Date created: 2017-02-15 11:37 AM | Last Updated: 2020-05-25 06:11 PM

Identifier: DOI 10.17605/OSF.IO/ZN79K

Category: Project

Description: We show that faces contain much more information about sexual orientation than can be perceived and interpreted by the human brain. We used deep neural networks to extract features from 35,326 facial images. These features were entered into a logistic regression aimed at classifying sexual orientation. Given a single facial image, a classifier could correctly distinguish between gay and heterosexual men in 81% of cases, and in 74% of cases for women. Human judges achieved much lower accuracy: 61% for men and 54% for women. The accuracy of the algorithm increased to 91% and 83%, respectively, given five facial images per person. Facial features employed by the classifier included both fixed (e.g., nose shape) and transient facial features (e.g., grooming style). Consistent with the prenatal hormone theory of sexual orientation, gay men and women tended to have gender-atypical facial morphology, expression, and grooming styles. Prediction models aimed at gender alone allowed for detecting gay males with 57% accuracy and gay females with 58% accuracy. Those findings advance our understanding of the origins of sexual orientation and the limits of human perception. Additionally, given that companies and governments are increasingly using computer vision algorithms to detect people's intimate traits, our findings expose a threat to the privacy and safety of gay men and women.

“We show that faces contain much more information about sexual orientation than can be perceived and interpreted by the human brain… Given a single facial image, a classifier could correctly distinguish between gay and heterosexual men in 81% of cases, and in 74% of cases for women. … Consistent with the prenatal hormone theory of sexual orientation, gay men and women tended to have gender-atypical facial morphology, expression, and grooming styles … our findings expose a threat to the privacy and safety of gay men and women.”

Wang & Kosinski 2017

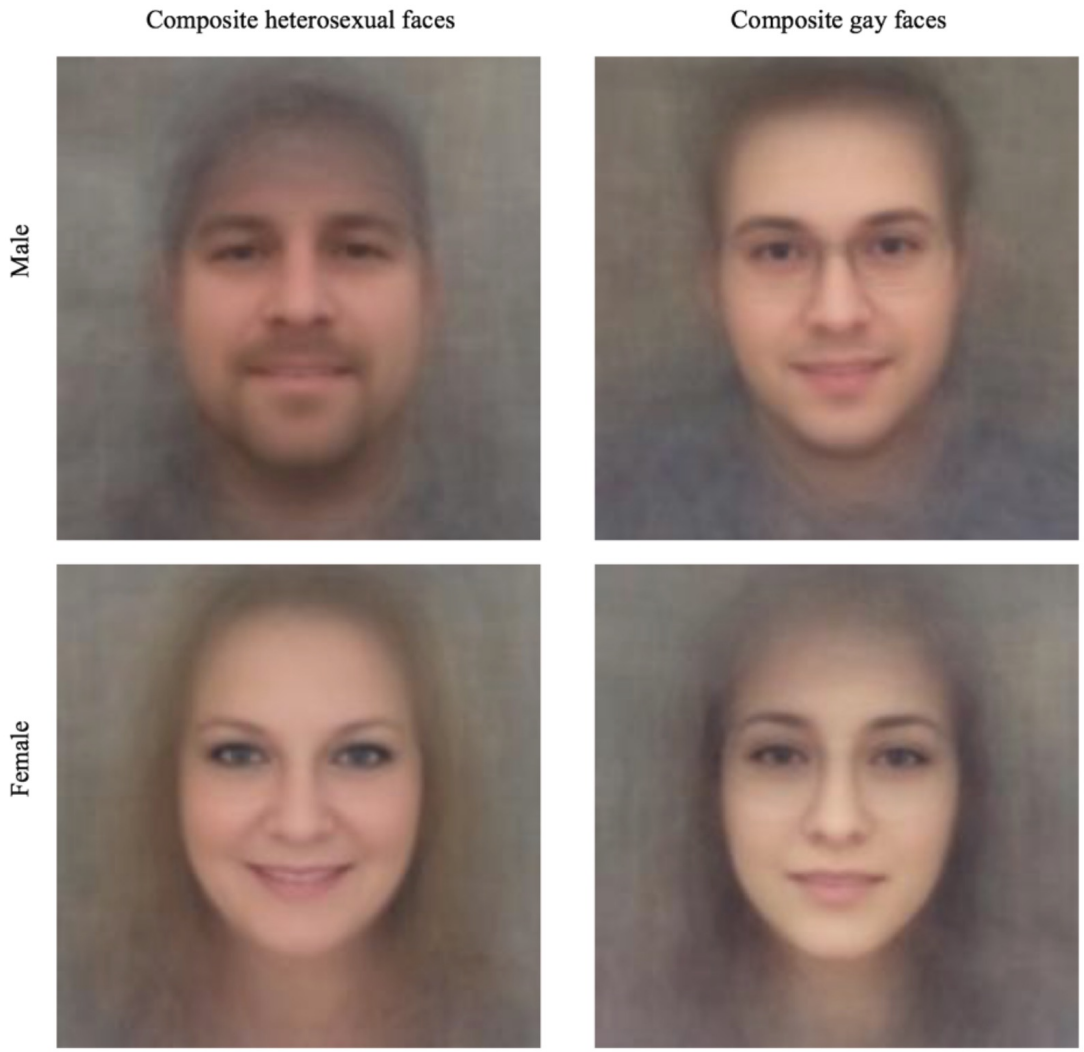
More questions

- **Does the application align with your values?**
- Does the task specification / evaluation metric reflect the things you care about?
- For recognition:
  - **Does the collected training / test set match your true distribution?**
- Are the algorithm's errors biased?
- **Are you being honest in public descriptions of your results?**
- Is the accuracy/correctness sufficient for public release?

Slide credit: Bharath Hariharan

Answers

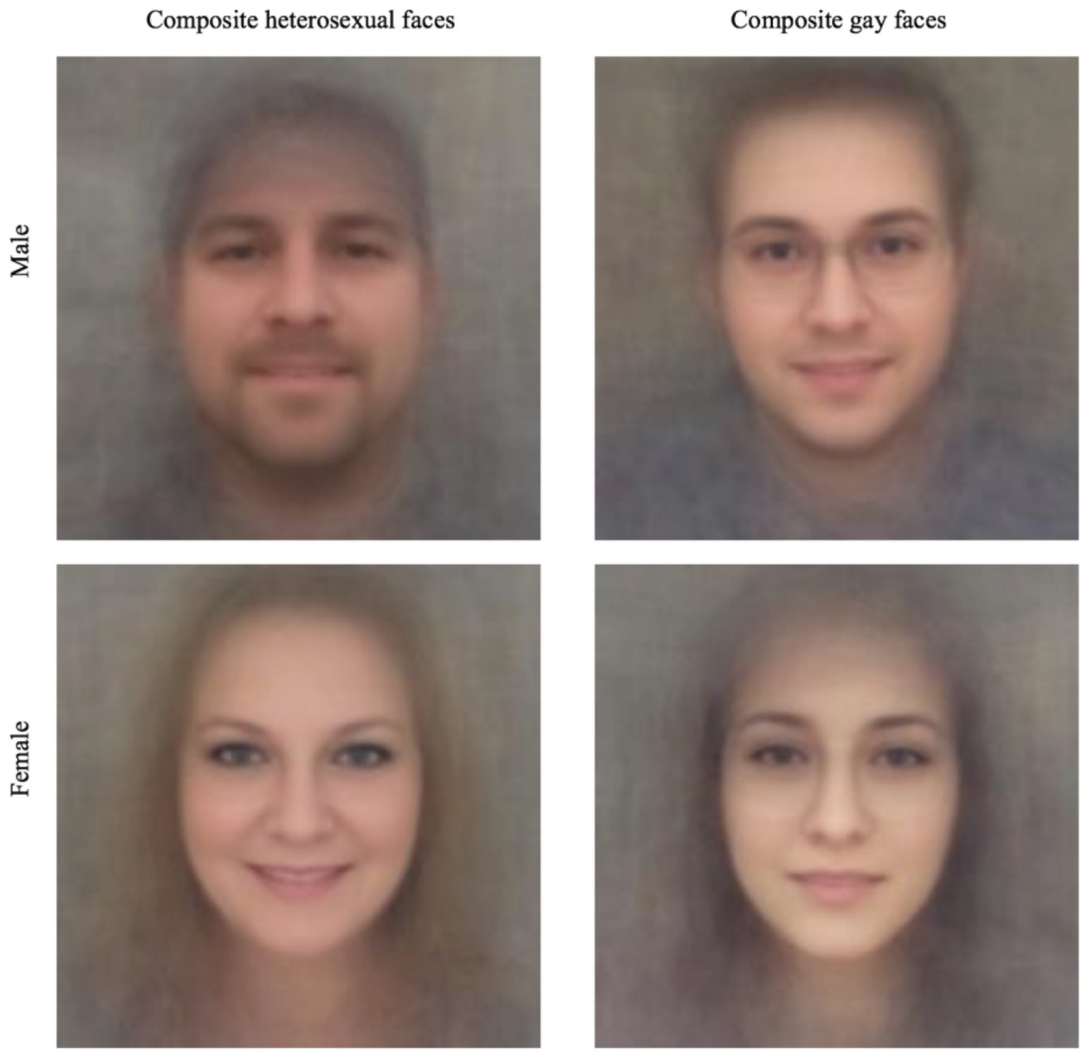
- Training / test set?
  - 35,326 images from public profiles on a US dating website
- "average" images of straight/gay people:
- Question:
  - Are differences caused by actual differences in faces
  - Or how people choose to present themselves in dating websites?



Slide credit: Bharath Hariharan

Answers

- Goal: raise privacy concerns.
- Side-effects?
  - Reinforces potentially harmful stereotypes
  - Provides ostensibly “objective” criteria for discrimination



Slide credit: Bharath Hariharan

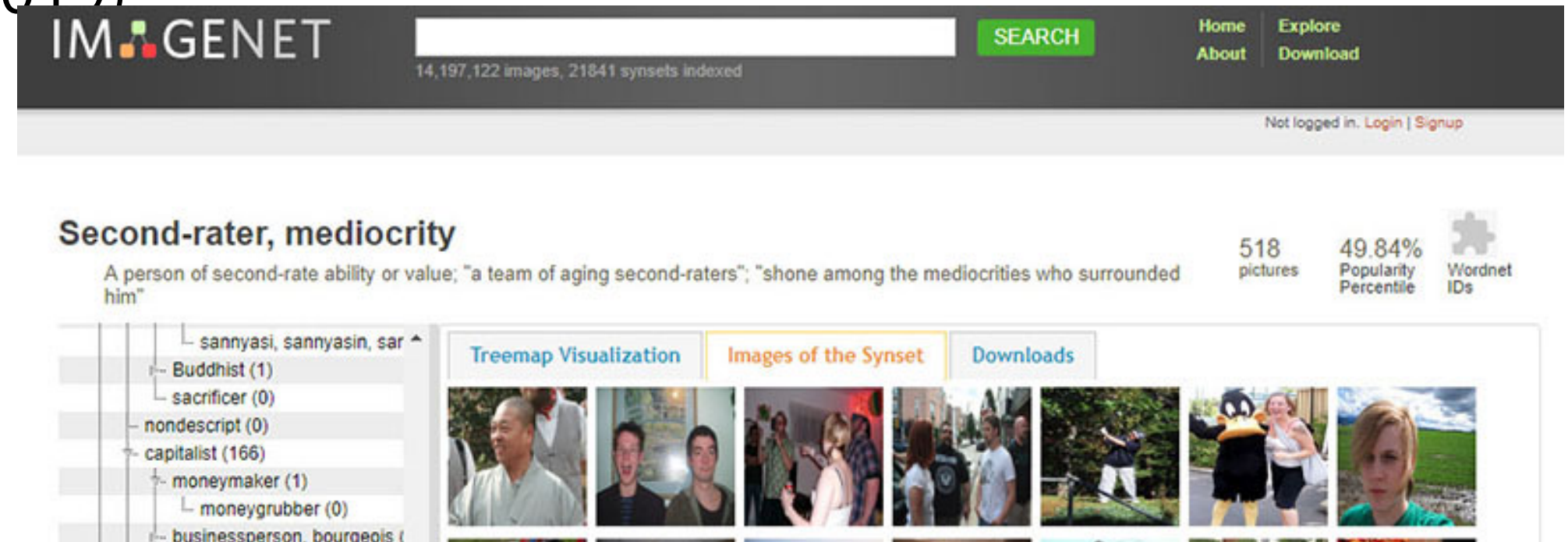
Do algorithms reveal sexual orientation or just expose our stereotypes?
Blaise Agüera y Arcas, Alexander Todorov and Margaret Mitchell
https://medium.com/@blaisea/do-algorithms-reveal-sexual-orientation-or-just-expose-our-stereotypes-d998fafdf477

Datasets – Potential Issues

- Licensing and ownership of data
- Consent of photographer and people being photographed
- Offensive content
- Bias and underrepresentation
  - Including amplifying bias
- Unintended downstream uses of data

Case study – ImageNet

- Serious issues with the People subcategory
  - Offensive content, non-visual categories
- Pointed out by https://excavating.ai/ (Crawford & Paglen, 2019)



Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy

Kaiyu Yang, **Klint Qinami**, **Li Fei-Fei**, **Jia Deng**, **Olga Russakovsky**

Abstract

Computer vision technology is being used by many but remains representative of only a few. People have reported misbehavior of computer vision models, including offensive prediction results and lower performance for underrepresented groups. Current computer vision models are typically developed using datasets consisting of manually annotated images or videos; the data and label distributions in these datasets are critical to the models' behavior. In this paper, we examine ImageNet, a large-scale ontology of images that has spurred the development of many modern computer vision methods. We consider three key factors within the `person` subtree of ImageNet that may lead to problematic behavior in downstream computer vision technology: (1) the stagnant concept vocabulary of WordNet, (2) the attempt at exhaustive illustration of all categories with images, and (3) the inequality of representation in the images within concepts. We seek to illuminate the root causes of these concerns and take the first steps to mitigate them constructively.

Case study – Microsoft Celeb

- Microsoft Celeb (MS-Celeb-1M): dataset of 10 million face images harvested from the Internet for the purpose of developing face recognition technologies.

- From http://exposing.ai: “While the majority of people in this dataset are American and British actors, the exploitative use of the term ‘celebrity’ extends far beyond Hollywood. Many of the names in the MS Celeb face recognition dataset are merely people who must maintain an online presence for their professional lives: journalists, artists, musicians, activists, policy makers, writers, and academics. Many people in the target list are even vocal critics of the very technology Microsoft is using their name and biometric information to build.”

Case study – Microsoft Celeb

- Microsoft Celeb taken down May 2019
- However, dataset still can be found online
- Case brings up questions of consent and privacy of individuals in a dataset, as well as uses of large-scale face recognition and “runaway datasets”

Some “sunsetted” datasets

- Microsoft Celeb (MS-Celeb-1M)
- ImageNet (partial – people category)
- MIT Tiny Images
- MegaFace
- Duke MTMC Dataset

- See https://exposing.ai/datasets/ for more information
- Additional reference:
  - *Large image datasets: A pyrrhic win for computer vision?* Vinay Uday Prabhu & Abeba Birhane. https://arxiv.org/abs/2006.16923

Datasheets for Datasets

"The ML community currently has no standardized process for documenting datasets, which can lead to severe consequences in high-stakes domains. To address this gap, we propose datasheets for datasets. In the electronics industry, every component, no matter how simple or complex, is accompanied with a datasheet that describes its operating characteristics, test results, recommended uses, and other information. By analogy, we propose that every dataset be accompanied with a datasheet that documents its motivation, composition, collection process, recommended uses, and so on."

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, Kate Crawford. 2018

DeepFakes

https://www.theverge.com/2021/3/5/22314980/tom-cruise-deepfake-tiktok-videos-ai-impersonator-chris-ume-miles-fisher

DeepFakes

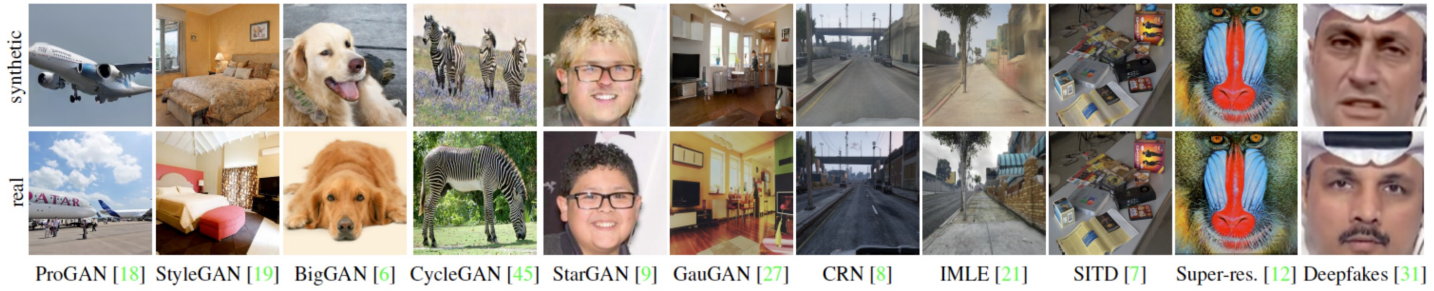
DeepFakes

- Active research on both better and better image/video generation and detection of fake images
- Representative work:

CNN-generated images are surprisingly easy to spot...for now

Sheng-Yu Wang[1] Oliver Wang[2] Richard Zhang[2] Andrew Owens[1] Alexei A. Efros[1]
[1]UC Berkeley [2]Adobe Research

Code [GitHub] CVPR 2020 (Oral) [Paper]



Are CNN-generated images hard to distinguish from real images? We show that a classifier trained to detect images generated by only one CNN (ProGAN, far left) can detect those generated by many other models (remaining columns).

https://peterwang512.github.io/CNNDetection/

Text-to-image models

- Often trained on datasets that contain copyrighted material

ARTIFICIAL INTELLIGENCE / TECH / LAW

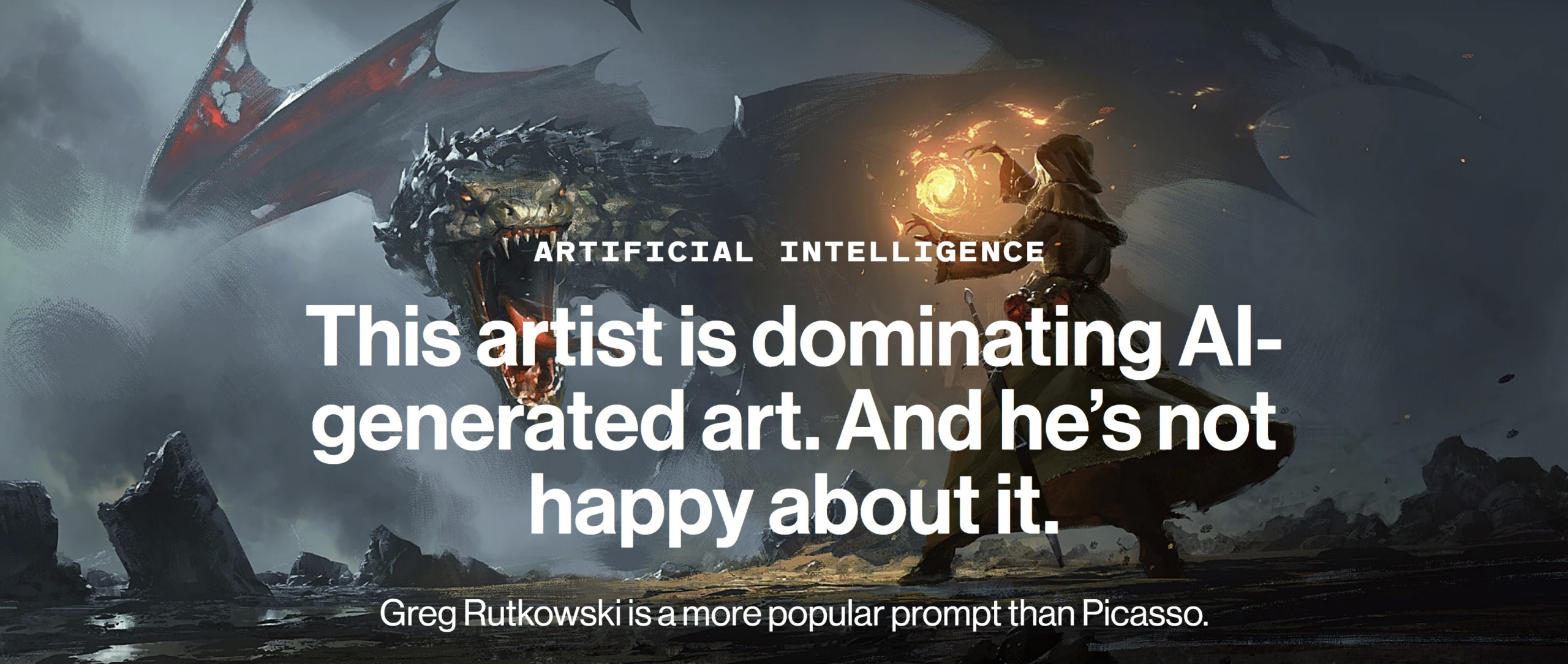
Getty Images is suing the creators of AI art tool Stable Diffusion for scraping its content

/ Getty Images claims Stability AI ‘unlawfully’ scraped millions of images from its site. It’s a significant escalation in the developing legal battles between generative AI firms and content creators.

By JAMES VINCENT

Jan 17, 2023, 5:30 AM EST | 18 Comments / 18 New

https://www.theverge.com/2023/1/17/23558516/ai-art-copyright-stable-diffusion-getty-images-lawsuit



"Dragon Cave"
GREG RUTKOWSKI

Example Text-to-image prompt: "Wizard with sword and a glowing orb of magic fire fights a fierce dragon Greg Rutkowski,"

https://www.technologyreview.com/2022/09/16/1059598/this-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/

A new AI draws delightful and not-so-delightful images

OpenAI's DALL-E 2 is incredible at turning text into images. It also highlights the problem of AI bias — and the need to change incentives in the industry.

By Sigal Samuel | Apr 14, 2022, 8:00am EDT

Generated images of lawyers

Generated images of flight attendants

https://www.vox.com/future-perfect/23023538/ai-dalle-2-openai-bias-gpt-3-incentives

Some tools

- Policy and regulation
  - e.g., a number of cities have banned the use of face recognition by law enforcement
- Transparency
  - e.g., studies on bias in face recognition have led to reforms by tech companies themselves
  - e.g., datasheets can help downstream users of datasets
- Awareness (when you conceive of or build a technology, be aware of the questions we've discussed)

Questions?