LET'S FACE IT: THE CHALLENGE OF PROTECTING IDENTITIES ON SOCIAL MEDIA

Shinjani Agnihotri

INTRODUCTION

Facial recognition technology ("FRT") has been a lightning rod for debate in the public space. The technology involves the creation of a template of human faces to identify the unique facial structure and tagging it to connect that face with the full record of other information provided by the users.[1] The proponents of this technology advocate for its larger benefits including surveillance in the wake of rising extremist views in democracies, while the opposing side has expressed their well-founded concerns for its underlying issues regarding algorithmic bias and privacy. FRT has been banned in public spaces in some US cities, including San Francisco, Boston, and Portland. California has passed legislation prohibiting the use of FRT in police body cameras for three years beginning January 1, 2020.[2]

FRT is used in everyday activities such as unlocking phones and paying for purchases. There are various open-source algorithms that detect facial features which makes it accessible for application ("app") developers. Its use has been expanded from video cameras in smart homes for identifying visitors to using it in smartphones' cameras for its auto-focus feature and video stabilization. Even Cloud photo storage creates albums and themed slideshows by grouping faces it recognizes from the images captured by users. Many apps, including social media filters, use face analysis to create effects like artificial aging and animating facial features. Self-improvement and beauty forecasting apps, horoscopes, and ethnicity detection also use facial scans to generate recommendations. The development of FRTs helps form larger systems for tracking populations, ranking clients, and developing suspect pools for investigations.

However, how is the dataset created that is used to train the Artificial Intelligence algorithm ("algorithm")? From where are the images taken to create said datasets? Algorithms are not innately capable of recognizing human faces, there has to be a process of feeding the data into them. So then, whose pictures are used to train these algorithms and were they informed of such use? These are some of the underlying issues of developing FRT

[1] Brenda Leong, *Facial recognition and the future of privacy: I always feel like … somebody's watching me*, 75(3) Bulletin of the Atomic Scientists 109, 110 (Apr. 26, 2019), https://doi.org/10.1080/00963402.2019.1604886.

[2] Gibson Dunn, 2019 *Artificial Intelligence and Automated Systems Annual Legal Review* 17 (Feb. 11, 2020), available at https://www.gibsondunn.com/wp-content/uploads/2020/02/2019-artificial-intelligence-and-automated-systems-annual-legal-review.pdf.

which are explored further in this paper. The scope of social media platforms ("platforms") includes those platforms where users can upload their images.

PRIVACY CONCERN

Data Privacy is the right of a person to choose whether, how, and to what extent information about themselves, particularly sensitive and confidential, is communicated to others.[3] This right is in relation to the possession of information by organizations or third-party, including facial and biometric information stored in databases.[4] As technology advances, creating boundaries in today's societal environment is by no means an easy task. The purpose of creating regulations to protect privacy is to develop a framework where the release, sharing, and use of online data is transparent so that privacy in our daily lives can be managed.[5] While there is a considerable debate when it comes to government employing methods like FRT for surveillance purposes, any unlawful collection of data by private stakeholders has always been condemned.

Online identities of individuals can be divided into three categories: first, "transactional identity" which is referred to a set of data that enables an individual to engage in a transactional relationship with an organization, such as a bank, government, or insurance company. Second, "social identity" is the sum of an individual's online data, such as posts, images posted on social media, location (check-ins), and frequency of posting. Finally, a "professional identity" is a person's collection of skills, competencies, and work experience that has been carefully curated for business and job-related purposes.[6] Most Internet users are unaware of the relationship between these three identities, or how their personal digital information can be used by employers and governments.

Harvesting data without consent is major privacy violation and can lead to adverse consequences. During mid-March of 2018, Cambridge Analytica was exposed for its extra-judicial dealings with the Trump campaign, in which the company harvested more than 87 million Facebook profiles without consent/legal justification.[7] This allowed Cambridge Analytica to create an algorithm that skewed news results in

---

[3] *Information Privacy*, Black's Law Dictionary (9th ed. 2009).

[4] Marcus Smith et al., *Facial Recognition and Privacy Rights*, Biometric Identification, Law and Ethics 30 (2021), https://link.springer.com/chapter/10.1007/978-3-030-90256-8_2.

[5] Sanjay Sharma, DATA PRIVACY AND GDPR HANDBOOK 5 (2020).

[6] Liam Bullingham et al., *The Presentation of Self in the Online World': Goffman and the Study of Online Identities*, 39 Journal of Information Science (2013).

[7] Jim Isaak, *User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection*, IEEE Computer Society 56 (2018).

Facebook users' news feeds. According to critics, the move was not only illegal, but it also had a significant impact on the outcome of the US election.[8]

In the modern-day, users expose their faces to the risk of automated scrutiny by consuming services of a platform in exchange for their data (both personal and social). The biometric data regarding social identity (facial images) is linked to other identities for building larger data sets, and then it is fed into algorithms for their development. Biometrics are immutable, easily accessible, individuating, and can be highly prejudicial. There is lucrative advantage for creation of new databases as the commercial value of FRT is derived from its ability to link the data collected and shared from third parties. Third parties can create their own database using images available on users' accounts on a platform and transform those databases into full-fledged identification systems capable of identifying individuals on a larger scale.[9]

Regulations from a policy perspective have failed to keep-up with the everyday usage of facial biometrics on platforms. Take Clearview AI, for example, it created its facial recognition database by scraping images of people from the internet, especially from social media. The company filed for a patent application as to how it acquires images using a "web crawler," with the caveat that online photos associated with an account may help to create additional records of facial recognition data points, which its machine learning algorithm can then use to find and identify matches.[10] It can identify people even when they are in the background and does not require photos of people looking directly at the camera.[11]

The American Civil Liberties Union filed a complaint against the company for violation of the Illinois Biometric Information Privacy Act (BIPA), by illegally collecting and storing data of citizens without their knowledge and consent.[12]

---

[8] H. Akın Ünver, *Politics of Digital Surveillance, National Security and Privacy*, Centre for Economics and Foreign Policy Studies, 2 (2018), http://www.jstor.org/stable/resrep17009.

[9] Amba Kak, ed., *Regulating Biometrics: Global Approaches and Urgent Questions*, AI Now Institute, 8 (2020), https://ainowinstitute.org/regulatingbiometrics.pdf.

[10] Emma Roth, *Clearview AI is closer to getting a US patent for its facial recognition technology*, The Verge (Dec 5, 2021), https://www.theverge.com/2021/12/5/22819097/clearview-ai-facial-recognition-patent.

[11] Elizabeth A. Rowe, *Regulating Facial Recognition Technology in the Private Sector*, 24 Stan. Tech. L. Rev. 8, 1-54 (2020), https://www-cdn.law.stanford.edu/wp-content/uploads/2020/12/Rowe-FINAL-Facial-Recognition.pdf.

[12] ACLU v. Clearview AI, No. 2020 CH 04353 (Ill. Cir. Ct., Cook City), https://www.aclu.org/sites/default/files/field_document/2020.05.28_aclu-clearview_complaint_file_stamped.pdf.

Subsequently, the company stopped the sale of its technology to private companies and non-law enforcement agencies. While the images are protected under BIPA, it is debatable whether scans of such photos are also protected.[13] The company has been facing scrutiny in other jurisdictions such as the UK, Australia, and Canada. However, it filed for a patent for its "search engine for faces" in the US,[14] which has been granted a notice of allowance: this implies that the patent will be awarded after payment of administrative fees.[15] This is not the first time that US has encouraged FRT without considering its long-term implications.[16]

Clearview AI is a direct example of a third-party using an automated web crawler to identify, collect and verify pictures of users. However, images uploaded by users on such platforms can be used by the platform itself to create their own database, which is equally violative of users' privacy. The whole justification stems from implied consent provided by users while setting up their account on the platform but is it umbrella consent to cover all kinds of usage.

DEEPFAKES ISSUE

Another cause for concern of publicly available images on platforms is the creation and rapid dissemination of DeepFakes, which are high-quality tampered videos where the face of one person is swapped onto another using neural networks.[17] Large amounts of synthetically generated DeepFake videos appear in the online sphere as a result of open-source software and apps for face swapping. This poses a significant technical challenge for the detection and filtering of such content. Public scandals involving the swapping of celebrities on pornographic content are just the tip of the iceberg, it can also be exploited to cause political tensions by creating false videos of world leaders.[18] The potential threat of this technology affecting the masses is significant. From destroying the reputation of an individual to breaching the security of e-transactions, the

[13] Elizabeth A. Rowe, *supra* note 11, at 26.
[14] U.S. Publ. No. 0042527 (filed Aug. 7, 2020).
[15] Alexandra S. Levine, *Clearview AI on track to win U.S. patent for facial recognition technology*, Politico (Apr 12, 2021), https://www.politico.com/news/2021/12/04/clearview-ai-facial-recognition-523735.
[16] *Face Detection and Recognition*, U.S. Patent No. 9639740 (filed Nov. 12, 2013).
[17] Pavel Korshunov et al., *DeepFakes: a New Threat to Face Recognition? Assessment and Detection*, 1 (2018), https://arxiv.org/pdf/1812.08685.pdf.
[18] Thanh Thi Nguyen et al., *Deep Learning for Deepfakes Creation and Detection: A Survey*, 1-2 (2019), https://arxiv.org/pdf/1909.11573.pdf.

ramifications are that digital identities of individuals are jeopardized.

The EU Agency for Cybersecurity has taken cognizance of this issue and published its report on remote identity proofing so as to minimize identification threats.[19] One of the major face presentation attacks identified by the report is DeepFakes through which attackers can create synthetic video realistically representing someone else by having access to a dataset containing images of their target.[20] The report identified social media as a good source of data for the purpose of such attacks. With open-source software that can construct DeepFakes from low-resolution images, the threat of misusing photos taken from social media only amplifies.[21]

INTELLECTUAL PROPERTY ANGLE: COPYRIGHT DEFENSE

Copyright is a unique intellectual property right since this right is vested with the author from the moment a copyrightable subject matter is created. The users are vested with copyright on images posted on social media. Thus, as per copyright law, an express authorization is required from each individual user before using their pictures to train the algorithm. However, technologies owe their growth to the defense of the "fair use" doctrine pursuant to which the subsequent user will not be bound by the mandate of taking express authorization from the original author if the subsequent use is "transformative" in nature. For that, such use must be productive, employed in a different manner, or for a different purpose from the original subject matter.[22]

Two prominent cases in this sphere are *Kelly* v. *Arriba Soft Corp.* and *Perfect 10, Inc.* v. *Amazon.com, Inc*. In *Kelly* v. *Arriba Soft Corp.*,[23] Arriba operated a search engine that returned thumbnail-sized images in response to search queries. When Arriba's crawler came across an image, it would download full-size copies to Arriba's servers. The images were then

---

[19] *Remote Identity Proofing: Attacks & Countermeasures, European Union Agency for Cybersecurity*, 3 (2022) https://www.enisa.europa.eu/publications/remote-identity-proofing-attacks-countermeasures.

[20] Press Release, *Beware of Digital ID attacks: your face can be spoofed!*, (Jan. 30, 2022), https://www.enisa.europa.eu/news/enisa-news/beware-of-digital-id-attacks-your-face-can-be-spoofed.

21 Alexandros Lattas et al., *AvatarMe: Realistically Renderable 3D Facial Reconstruction "in-the-wild", in* Conference on Computer Vision and Pattern Recognition 767, 760-769 (2020), https://openaccess.thecvf.com/content_CVPR_2020/papers/Lattas_AvatarMe_Realistically_Renderable_3D_Facial_Reconstruction_In-the-Wild_CVPR_2020_paper.pdf.

[22] Pierre N. Leval, *Toward a Fair Use Standard*, 3 Harvard L. Rev. 1111 (1990), https://www.jstor.org/stable/1341457.

[23] Kelly v. Arriba Soft Corp, 336 F.3d 811 (9th Cir. 2003).

reduced to thumbnail size, the full-size copies were deleted, and the thumbnails were featured in Arriba's search results. Similarly in *Perfect 10, Inc. v. Amazon.com, Inc.*,[24] Perfect10 availed services of Google image search which created, stored, and displayed thumbnails of the original images.

For both the decisions, the doctrine of non-expressive use was relied on to conclude that the search engines were mere tools and not mediums of conveying the original expression.[25] The original images were uploaded for their aesthetic value as an artistic work whereas the search engines were indexing their thumbnail version to direct the users to the original image. It was unlikely that the users would consume thumbnail versions of the images for aesthetic purposes and thus, the use of images by search engines was considered to be transformative.

Whether transformative use can be applied to machine learning is yet to be evaluated. It could be contended that the use of photographs of users on a platform is covered under the fair use exception since the images are not used for their aesthetic value and their purpose has been transformed to modify them as datasets for training the AI. Another contention could be regarding consent provided by users by agreeing to the terms of service of the platform. This may amount to authorization to use images of users posted on the platform by virtue of making an account on that platform. However, this reasoning is not infallible and could be refuted by "Notice-and-Choice." This is based on the legal doctrine that, as long as the platform discloses that information regarding data collection such as the extent, manner and purpose of data processing and the users are provided with discretion *ex-ante*, then the subsequent choice is informed and valid. Since with every platform, there is no choice on the part of users to disallow that platform from using their images, such "all-or-nothing" consent is not proper, and thus, invalid.

CASE STUDY: SCOPE OF CONSENT IN FACEBOOK

FRT is sophisticated in its application, mainly due to the lack of strong legislation regarding its extent and depth. As a result of this democratic and legal deficit, citizens are forced to look after their own privacy defenses, generating significant momentum in favor of social media sites with their ambiguous and ever-changing policies of secrecy. However, with FRT becoming mainstream, fewer options are available. Where 69% of adults use Facebook and seven out of ten users use it on a daily basis, the need to be available and updated on each platform only

[24] Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146 (9th Cir. 2007).
[25] Benjamin L. W. Sobel, *Artificial Intelligence's Fair Use Crisis*, Columbia Journal of Law & the Arts, Forthcoming 8, 1-49 (2017).

exacerbates the situation.[26] Using social media to connect with friends, family, and the rest of the world has quickly become a daily and necessary experience.

Facebook operated an algorithm that identified people in an image posted on the platform and notified each of them, giving them an option to tag themselves in that photograph. Although the facility has been shut down vide an official announcement by the company,[27] the manner of collection employed by the platform requires scrutiny. A similar challenge has been raised in a Texas court wherein the company is sued for illegally obtaining Texans' data in violation of Texas biometric laws.[28] While obtaining images of Facebook users, images of people who are not on Facebook were also taken, leaving such people with no remedy against this unauthorized exploitation.

In its terms of service, Facebook requires users to provide legal permission (known as 'license') to use their images. The scope of this license is to provide better services and enhance the experience of users online. Under the regimes of General Data Protection Regulation (GDPR) in EU[29] and the now defunct EU-US Privacy Shield,[30] the element of *notice* requires platforms to inform the users regarding the manner of data processing and the *purpose limitation* requires that data should be collected only to the extent necessary for the original purpose cited.[31] The purpose cited by Facebook was to facilitate its services to the users but the users were not expressly informed that their images were being processed by the platform to become part of the dataset for its algorithm.

Additionally, there is no security policy preventing someone else from taking snapshots of images. While Facebook users can opt for the safety feature of 'locking' their profile, it is not infallible. It only takes a methodological approach to procure the image. Platforms are required to observe due diligence when

---

[26] Brooke Auxier et al., *Social Media Use in 2021*, Pew Research Centre (Apr. 7, 2021), https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.

[27] Jerome Pesenti, *An Update on Our Use of Face Recognition*, Facebook (Nov. 2, 2021), https://about.fb.com/news/2021/11/update-on-use-of-face-recognition/.

[28] The State of Texas v. Meta Platforms, Inc., https://texasattorneygeneral.gov/sites/default/files/images/child-support/State%20of%20Texas%20v.%20Meta%20Platforms%20Inc..pdf.

[29] Regulation 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119/1) 1 (EU).

[30] Commission Implementing Decision (EU) 2016/1250 of 12 July 2016 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the EU-US Privacy Shield (notified under document C(2016) 4176) (Text with EEA relevance), 2016 O.J. (L207/1).

[31] Sanjay Sharma, *supra* note 5, at 9.

it comes to data protection, but Facebook's terms of service contain limitations to that liability, which states:

> *"…Our Products, however, are provided "as is", and we make no guarantees that they will always be safe, secure or error-free, or that they will function without disruptions, delays or imperfections. To the extent permitted by law, we also DISCLAIM ALL WARRANTIES, WHETHER EXPRESS OR IMPLIED, INCLUDING THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. We do not control or direct what people and others do or say, and we are not responsible for their actions or conduct…*
>
> *Accordingly, our liability shall be limited to the fullest extent permitted by applicable law… even if we have been advised of the possibility of such damages."*[32]

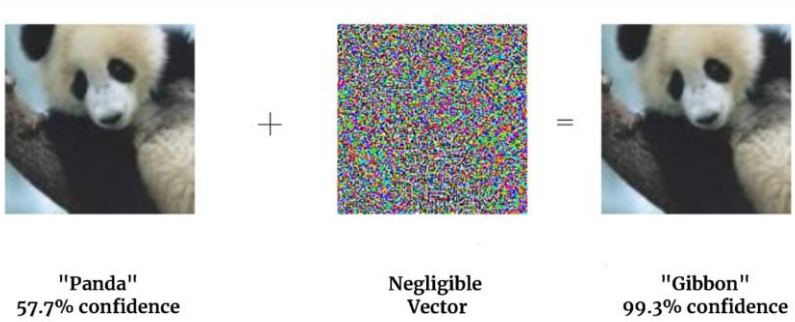
The clause disclaims all the liability that may arise out of their products (their own database) or third-party behavior (such as Clearview AI) where images can be procured by someone else. Once the "digital dignity" of users is compromised, their biometric identification can be exploited for various purposes, from impersonation to unauthorized tracking. While the consent of users may be valid when it comes to retention of data by Facebook for *limited purposes*, the transformation of images into literal datasets goes beyond the consent provided by individuals. This added with the fact that Facebook holds no accountability for the grave consequences that may arise for the users, calls for state intervention to form requisite regulatory mechanisms.

PROPOSED SOLUTION

Undoubtedly, FRT poses a number of privacy issues and requires oversight mechanisms for pre-emptive measures instead of damage-control measures. Regulations are intended to establish and monitor public safety safeguards. The solution in this case lies with technology itself. Arising out of a mistake, the concept of adversarial attack has posed a significant challenge for developers of FRT. Put simply, it is small perturbations to

[32] *Terms of Service*, Facebook, https://www.facebook.com/terms.php (last visited Feb 15, 2022).

inputs that do not change content for humans but cause a neural network to malfunction and misidentify the target.[33]



When starting with an image of a panda, a neural network identifies it as "panda" with 57.7% confidence. However, when the same is subjected to a negligible vector, the neural network misidentifies the same image as a gibbon with 99.3% confidence. Although the image has no visible difference for a human, the result of a neural network changes significantly with a greater degree of confidence for the wrong identification.[34]

The idea is to mandate platforms to create a tool that protects users from unauthorized facial recognition. A tool that can pre-process users' images in an adversarial manner before they are uploaded to social media. For third-party organizations collecting images for facial recognition, these pre-processed images would be useless. This will also protect images of users from being used by the platform without explicit authorization from the users. There have been past instances where technology was the best solution to technological problems. Anomaly detection to prevent online frauds could be one such example, where machine learning is applied in a series of online transactions for detecting intrusion and alerting the authorities of any unusual behaviors/fraud. Another example could be strengthening security measures such as two-factor authentication for online transactions to ensure that the account holder is making a transaction. Similarly, this issue also has a techno-legal solution.

CONCLUSION

With so much uncertainty as to the development of FRT in the future and potential complications it might pose on an international level since the same could be used for surveillance of users in foreign jurisdictions as well, it is imperative to bring pre-emptive regulations. Currently, various platforms have kept consent to use pictures of users as "necessary consent" and users

[33] Valeriia Cherepanova et al., *Lowkey: Leveraging Adversarial Attacks to Protect Social Media Users from Facial Recognition*, ICLR 2 (2021), https://arxiv.org/pdf/2101.07922.pdf.
[34] Ian J. Goodfellow et al., *Explaining and Harnessing Adversarial Examples*, ICLR 3 (2015), https://arxiv.org/pdf/1412.6572.pdf.

cannot avail services of the platform after declining the same. Adversarial attack provides an alternative to this "all-or-nothing consent" approach whereby the users can have the option to decline and still avail services of the platform. It is proposed that guidelines must be introduced for the states whereby platforms will have to provide an option for the users to deny licensing their images to platforms and pre-process their images. This will prevent both the platform and third parties from building their database. However, the platform can be permitted to retain those images for which informed consent must be obtained by the user.