BEING "SEEN" VS. "MIS-SEEN": TENSIONS BETWEEN PRIVACY AND FAIRNESS IN COMPUTER VISION

ALICE XIANG

ABSTRACT

The rise of AI technologies has been met with growing anxiety over the potential for AI to create mass surveillance systems and entrench societal biases. These concerns have led to calls for greater privacy protections and fairer, less biased algorithms. An under-appreciated tension, however, is that privacy protections and bias mitigation efforts can sometimes conflict in the context of AI. For example, one of the most famous and influential examples of algorithmic bias was identified in the landmark Gender Shades paper,[1] which found that major facial recognition systems were less accurate for women and individuals with deeper skin tones due to a lack of diversity in the training datasets. In an effort to remedy this issue, researchers at IBM created the Diversity in Faces (DiF) dataset,[2] which was initially met with a positive reception for being far more diverse than previous face image datasets.[3] DiF, however, soon became embroiled in controversy once journalists highlighted the fact that the dataset consisted of images from Flickr.[4] Although the Flickr images had Creative Commons licenses, plaintiffs argued that they had not consented to having their images used in facial recognition training datasets.[5] In part due to this controversy, IBM announced it would be discontinuing its facial recognition program.[6] Other companies that used the DiF dataset were also sued.[7] This example highlights the tension that AI technologies create between representation vs. surveillance, being "seen" vs. being "invisible." We want AI to "recognize" us, but we are uncomfortable with the idea of AI having access to personal data about us.

This tension is further amplified when the need for sensitive attribute data is considered. For example, in order to even discern whether a training dataset is diverse, we need a taxonomy of demographic categories, some notion of an ideal distribution across that taxonomy, and labels of these demographic categories. The methods that have emerged to address these necessities are

[1] Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, PROCEEDINGS OF MACHINE LEARNING RESEARCH 81:1–15, CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY (2018), http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

[2] Michele Merler et al., *Diversity in Faces* (2019), https://arxiv.org/pdf/1901.10436.pdf.

[3] Kyle Wiggers, *IBM Releases Diversity in Faces, a Dataset of Over 1 Million Annotations to Help Reduce Facial Recognition Bias*, VENTUREBEAT (Jan. 29, 2019), https://venturebeat.com/2019/01/29/ibm-releases-diversity-in-faces-a-dataset-of-over-1-million-annotations-to-help-reduce-facial-recognition-bias/.

[4] Taylor Shankland, *IBM Stirs Controversy by Using Flickr Photos for AI Facial Recognition*, CNET (Mar. 13, 2019), https://www.cnet.com/news/ibm-stirs-controversy-by-sharing-photos-for-ai-facial-recognition/.

[5] Taylor Hatmaker, Lawsuits Allege Microsoft, Amazon and Google Violated Illinois Facial Recognition Privacy Law, TECH CRUNCH (Jul. 15, 2020), https://techcrunch.com/2020/07/15/facial-recognition-lawsuit-vance-janecyk-bipa/.

[6] Nicolas Rivero, *The Influential Project that Sparked the End of IBM's Facial Recognition Program*, QUARTZ (June 10, 2020), https://qz.com/1866848/why-ibm-abandoned-its-facial-recognition-program/

[7] https://techcrunch.com/2020/07/15/facial-recognition-lawsuit-vance-janecyk-bipa/

often discomfiting and raise further privacy concerns. In designing DiF, the researchers did not have a variable for race, so they used various computational methods to derive labels for different facial features to approximate differences across race, including metrics for skin color and craniofacial areas.[8] While these features were used in an effort to ensure racial diversity without access to direct data on race, these approaches do not capture the sociological nature of demographic labels and could be misused, as we have seen in the pseudoscience of physiognomy, which focuses on quantifying physical differences across races.[9] Other attempts at creating diverse face image datasets, like FairFace,[10] approach the challenge by having Mechanical Turkers (MTurkers) guess people's demographic attributes. If at least two Turkers agree, then the label is considered ground truth; if there is no agreement, the image is discarded. This approach is concerning in that it relies on the ability of MTurkers to accurately assess people's demographic attributes, and it discards the images of people who might not fit neatly in the demographic taxonomy. This could, for example, lead to fewer multiracial, non-binary, or transgender individuals being represented in the data. Designing a taxonomy for demographic classification often relies on stereotypes and can impose and perpetuate existing power structures.

Existing privacy law addresses this issue primarily by erring on the side of hiding people's data unless there is explicit informed consent. In fact, privacy law and anti-discrimination law are often viewed as symbiotic,[11] under the assumption that preventing companies from collecting protected attribute data helps to prevent discrimination. It is intuitive to think that not being "seen" by AI is preferable—that being under-represented in training data might somehow allow one to evade mass surveillance. As we have seen in the policing context, however, just because facial recognition technologies do not work as well at identifying people of color has not meant that they have not been used to surveil these communities and deprive individuals of their liberty. In 2019, for example, Nijeer Parks, a Black man, was arrested due to a faulty facial recognition match.[12] He spent ten days in jail and paid around $5,000 to defend himself before the case was dismissed for lack of evidence. Thus, not being "seen" by AI does not protect against being "mis-seen."

The first contribution of this Article is to characterize this tension between privacy and fairness in the context of algorithmic bias mitigation for computer vision technologies. In particular, this Article argues that the irreducible paradox underlying current efforts to design less biased algorithms is the simultaneous desire to be "unseen" yet not "mis-seen" by AI. Second, the Article reviews the strategies that have been proposed for resolving this tension and evaluates their viability for adequately addressing the technical, operational, legal, and ethical challenges surfaced by this tension. These strategies include: using third-party trusted entities to collect data, using privacy-preserving techniques, obtaining informed consent, and creating regulatory mandates or government audits. Finally, this Article argues that solving this paradox

[8] *See supra* note 2 at 3.

[9] *See* Blaise Agüera y Arcas et al., *Physiognomy's New Clothes*, MEDIUM (May 6, 2017), https://medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a.

[10] Kimmo Kärkkäinen & Jungseock Joo, FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age (2019), https://arxiv.org/pdf/1908.04913.pdf.

[11] Jessica L. Roberts, *Protecting Privacy to Prevent Discrimination*, 56 WM. & MARY L. REV. 2097 (2015), https://scholarship.law.wm.edu/wmlr/vol56/iss6/4.

[12] Kashmir Hill, Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match, N.Y. TIMES (Dec. 29, 2020), https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html.

requires considering the importance of not being “mis-seen” by AI rather than simply being “unseen.” Detethering these notions (being seen, unseen, vs. mis-seen) can bring greater clarity around what rights relevant laws and policies should seek to protect. For example, this Article will examine whether, instead of just having regulations around what data should remain unseen by algorithmic systems, what the implications would be of a right not to be mis-seen by AI. Given that privacy and fairness are both extremely important objectives for ethical AI, addressing this tension head-on will be vital; approaches that rely purely on visibility or invisibility will likely fail to achieve either.

TABLE OF CONTENTS

Introduction 5
I: Definitions 7
II: Why Computer Vision? 10
III: Challenges to Algorithmic Bias Mitigation in Computer Vision 12
A. Why is Collecting Images with Informed Consent So Difficult? 14
IV: Privacy Laws 17
V: Harms of Being Seen vs. Mis-seen 19
A. Harms of Being Seen 19
B. Harms of Being Mis-seen 23
1. Lack of Protections Against Being Mis-Seen 25
VI: Approaches to Balancing Privacy and Bias Mitigation 26
A. Carve-Outs from Privacy Law 26
B. Participatory Design 28
C. Government or Trusted Third-Party Data Collection 29
D. Technological Advances 31
E. Right Against Being Mis-Seen 32
Conclusion 35

INTRODUCTION

Human-centric computer vision (HCCV),[13] including facial recognition, is one of the most controversial AI technologies in the U.S. and E.U. From a privacy perspective, the specter of mass surveillance, particularly by state actors, has led to significant criticism of the growing pervasiveness of the technology.[14] In addition, in recent years, there has been a growing awareness of the issues of bias in facial recognition technology. The highly influential Gender Shades paper showed that many of the major gender classification algorithms released by technology companies performed less well on women than men and less well on individuals with deeper skin tones than lighter skin tones.[15] Since then, subsequent studies, including one by the National Institute of Standards and Technology (NIST), a part of the U.S. Department of Commerce, have shown differences in performance on the basis of skin tone and gender for different computer vision systems.[16]

These findings have further fueled controversies around facial recognition technologies, and their real-world effects have been felt with cases of black men in the U.S. being wrongfully arrested due to faulty facial recognition matches.[17] Many jurisdictions have since put in place moratoriums on the usage of facial recognition by law enforcement and other public agencies.[18] This is notable given that while the growing pervasiveness of AI has led to substantial public discourse about the potential harms of such systems, facial recognition systems are the main ones to have been banned so far. Indeed, in the recent E.U. proposed AI regulations, one of main categories of prohibited technologies is the use of remote biometric identification (RBI) by law enforcement (with some narrow carve-outs), specifically targeting facial recognition.[19] Moreover, any RBI technology is considered to be high risk, and thus subject to extensive regulatory requirements around transparency and pre-deployment evaluation.[20]

[13] As will be discussed further in the Definitions section below, HCCV in this Article refers to computer vision systems that rely on images of humans for training and/or testing. This is a more specific subset of the "human-centered machine learning" models that Model Cards focus on, https://arxiv.org/pdf/1810.03993.pdf. HCCV is a more expansive term than Facial Processing Technologies (FPT), which encompasses "any task involving the identification and characterization of the face image of a human subject," https://arxiv.org/pdf/2102.00813.pdf. HCCV includes tasks involving human bodies and objects. HCCV can also be seen as any computer vision system relying on "people-centric" datasets, https://arxiv.org/pdf/2011.13583.pdf.

[14] See, e.g., https://edri.org/our-work/facial-recognition-document-pool/, https://www.amnesty.org/en/latest/news/2021/01/ban-dangerous-facial-recognition-technology-that-amplifies-racist-policing/, https://www.aclu.org/issues/privacy-technology/surveillance-technologies/face-recognition-technology, https://www.nature.com/articles/d41586-020-03188-2.

[15] http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

[16] https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf

[17] See, e.g., https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html, https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html, https://www.freep.com/story/news/local/michigan/detroit/2020/07/10/facial-recognition-detroit-michael-oliver-robert-williams/5392166002/.

[18] See, e.g., https://epic.org/state-policy/facialrecognition/ (listing moratoriums or bans in California and Massachusetts, https://slate.com/technology/2021/07/maine-facial-recognition-government-use-law.html (describing Maine's ban), https://kingcounty.gov/council/mainnews/2021/June/6-01-facial-recognition.aspx (describing King County's ban in Washington state).

[19] Title II, Article 5, 1(d), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

[20] Annex III, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

Despite the controversy around facial recognition and other HCCV technologies, their use continues to grow.[21] The North American market for facial recognition alone is expected to double by 2027.[22] Even though recent moratoriums on facial recognition technology suggest a strong discomfort with government use of the technology, the demand for private surveillance camera systems with this technology has continued to grow,[23] as has the use of this technology in everyday life. It is now common for people to open their phones with face verification, to automatically sort photos based on the people identified in the photos, and to apply filters to their faces using apps that map facial landmarks. Moreover, in Asia, where facial recognition has not faced as much controversy as in the U.S. or E.U.,[24] the technology is increasingly used not only by government authorities[25] but also for everyday verification purposes, such as payment[26] and entering establishments.[27]

Thus, while privacy and bias concerns about facial recognition technologies have manifested themselves in moratoriums in some jurisdictions, the growing use of the technology suggests that it is here to stay. The key question then for both policymakers and AI developers is how to address these privacy, bias, and other ethical concerns in contexts where the technology is in use now or might be in use in the future. Unfortunately, as this Article will discuss, addressing privacy and bias concerns in practice can be quite difficult—not only because each set of concerns entails addressing many sociotechnical challenges, but also because privacy and bias mitigation are often in tension in the algorithmic context. The goal of this Article is not to advocate for the increased or decreased usage of facial recognition or other HCCV technologies, but rather to characterize this tension between privacy and bias mitigation efforts in the computer vision context and to propose potential paths forward.

Section I lays out definitions for terms used throughout the Article. Section II explains what makes the computer vision context unique in terms of the privacy and fairness tensions it raises. Section III discusses current challenges to mitigating algorithmic bias in the computer vision context, focusing particularly on the difficulties with collecting large, diverse datasets with informed consent. Section IV discusses relevant privacy laws in the U.S. and E.U. Section

[21] https://www.prnewswire.com/news-releases/at-18-1-cagr-facial-recognition-market-size-expected-to-reach-usd-11604-5-million-by-2027-says-brandessence-market-research-301351001.html

[22] https://www.ft.com/content/f6a9548a-a235-414e-b5e5-3e262e386722

[23] See, e.g., https://www.techrepublic.com/article/demand-for-video-surveillance-cameras-expected-to-skyrocket/, https://www.theguardian.com/commentisfree/2021/may/18/amazon-ring-largest-civilian-surveillance-network-us, https://cybernews.com/privacy/the-rise-of-the-private-surveillance-industry/, https://www.cepro.com/security/global-video-surveillance-market-revenues-exceed-24b-2021/.

[24] https://arxiv.org/pdf/2008.07275.pdf. This is not to say there is no controversy around facial recognition in Asia. In fact, there is growing concern about biometric privacy in China. https://www.washingtonpost.com/world/facial-recognition-china-tech-data/2021/07/30/404c2e96-f049-11eb-81b2-9b7061a582d8_story.html, https://www.bbc.com/news/technology-50674909, https://www.wsj.com/articles/in-china-paying-with-your-face-is-hard-sell-11600597240. That said, the use of facial recognition is far more pervasive in China than in other countries. https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html

[25] See, e.g., https://www.nbcnews.com/tech/tech-news/singapore-facial-recognition-getting-woven-everyday-life-n1242945; https://www.npr.org/2021/01/05/953515627/facial-recognition-and-beyond-journalist-ventures-inside-chinas-surveillance-sta

[26] https://www.bbc.com/news/business-55748964, https://www.japantimes.co.jp/news/2021/08/02/business/tech/facial-recognition-tie-up/

[27] https://asia.nikkei.com/Business/Business-trends/Japan-in-race-with-China-for-facial-recognition-supremacy; https://www.reuters.com/article/us-health-coronavirus-japan-facial-recog/masks-no-obstacle-for-new-nec-facial-recognition-system-idUSKBN29C0JZ

V elaborates on the harms associated with being "seen" vs. "mis-seen." Section VI evaluates potential solutions for better balancing protections against being "seen" vs. "mis-seen."

I: DEFINITIONS

Throughout this Article, I will use the term "human-centric computer vision" (HCCV) to refer specifically to AI systems that rely on images of humans for their training data and test data.[28] Given the focus of this Article on privacy and fairness issues, these are the most relevant AI systems to examine. Of course, even if images of humans are not involved, there might still be legal issues with copyright, and privacy risks might still remain if photos taken inside people's homes are used. The main focus of the privacy analysis in this Article, however, is on protections around biometric information. The processing of biometric information is arguably the highest risk aspect of HCCV given the relative immutability of people's biometrics, and thus has been the primary target of regulatory protections.

The primary computer vision tasks motivating this piece are facial recognition, detection, verification, and classification, but I use the more expansive term of HCCV since many of my points also apply to body detection, pose estimation, and body recognition. Object detection and classification are also relevant insofar as developers use images of people and objects in order to train their models.

Although colloquially HCCV technologies are often generically referred to as "facial recognition technologies" (FRT), it is important to draw the distinction between HCCV and FRT. While HCCV encompasses all technologies whose development might confront biometric privacy laws, these laws are typically motivated by the desire to tackle FRT. In addressing the tensions between existing privacy laws and HCCV bias mitigation efforts, it is thus important to note that HCCV includes technologies, as enumerated below, that largely do not figure in policy conversations about biometric information privacy laws. Moreover, the legal and technical analysis of relevant harms and solutions differs to some extent based on the precise task, so in this Article, I will use more granular distinctions when relevant.

Face detection involves detecting whether a human face is in an image and, if so, drawing a bounding box around the face. This is one of the most frequently used face-related computer vision tasks and serves as the basis for the other face-related tasks (you must first detect a face before you can identify or analyze it). Face or body detection is often used to count people or to trigger a subsequent task. For example, an AI-assisted AC system for an office might only turn on if a human is detected as being in the room.

Face verification and recognition are related tasks for assessing who a person is. Face verification refers to a one-to-one comparison between a reference face and a new face. When unlocking a phone, a face verification algorithm is used to compare the face in front of the camera with the reference face for the owner of the phone. Facial recognition refers to one-to-many comparisons, when you have a new face you are comparing against a reference set of faces to identify which (if any) of the reference faces is a match. If police have an image of a suspect,

[28] See supra note [] for discussion of related terms in the existing literature.

they can run that image through a facial recognition system that compares the image to an existing database of driver's license photos to see if there is a match.

Face classification, also known as "facial analysis," refers to the task of automatically generating labels for a face. For example, the model might label faces as "male" or "female." This type of task can be fraught from an ethical perspective given concerns around how much information can be accurately discerned from someone's face. Gender classification has been criticized since, regardless of whether gender is viewed as purely a social construct or as a biological concept, it cannot be assessed purely based on a photo, especially if an individual is transgender or non-binary.[29] In addition, controversial technologies like emotion recognition and computer vision character/fitness assessments fall under this category. Research suggests that emotion recognition is largely unreliable because people's facial expressions do not directly reflect their emotions—e.g., you might smile through discomfort or sadness.[30] In addition, efforts to use face classification to identify who might be a better job candidate or who might have a propensity to criminal behavior have been highly criticized as pseudoscientific.[31] That said, facial analysis can also be used for more benign purposes, such as a "smile setting" on a camera that waits until everyone in the frame is smiling before taking a photo.[32]

Body detection/verification/recognition/analysis tasks are analogous to the face-related tasks above, except that the focus is on the entire body rather than the face. Body detection, for example, might be used by an autonomous vehicle to detect and avoid pedestrians. Pose estimation is also a common task in this grouping and is used to estimate the spatial locations of a person's joints to determine whether an individual is doing a certain activity. In a security context, the goal might be to detect whether someone is shoplifting or making rapid movements that might be dangerous. Such technologies are also commonly used for augmented reality or CGI experiences. Pose estimation does not necessarily involve identifying the person, but can be used for such purposes. Gait recognition—leveraging the patterns unique to each person's gait to identify an individual—is recognized as a form of biometric identification, which is subject to relevant biometric information privacy laws in the U.S. and E.U.[33]

Object detection and recognition is another major category of computer vision tasks. For example, a traffic camera might learn to detect and count the number of cars at an intersection. While such tasks might seem unrelated to HCCV, they are often trained using images featuring humans. For example, in the research community, the COCO dataset is one of the most commonly used datasets for object-related tasks.[34] This dataset features around two hundred thousand images with humans and objects labelled. Using images with humans can be helpful given that, in the real-world, we are often interested in detecting objects that humans are

[29] https://ironholds.org/resources/papers/agr_paper.pdf

[30] https://www.nature.com/articles/d41586-020-00507-5, https://www.ft.com/content/c0b03d1d-f72f-48a8-b342-b4a926109452

[31] https://medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a, https://www.bbc.com/news/technology-53165286, https://fortune.com/2021/01/19/hirevue-drops-facial-monitoring-amid-a-i-algorithm-audit/

[32] https://www.wsj.com/articles/SB120889435178135615

[33] EU Proposed AI regulation, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020XX1117%2801%29&qid=1627962106278, https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5 (CCPA), https://www.nysenate.gov/legislation/bills/2019/a6787 (NY), https://www.insideprivacy.com/eu-data-protection/uk-ico-publishes-new-guidance-on-special-category-data/, https://iapp.org/resources/article/biometrics/

[34] https://cocodataset.org/

interacting with. Training an object recognition model exclusively on images without humans might make it more difficult for the model to perform well in real-world contexts. In addition, often the goal is not object recognition in isolation but rather to combine human detection with object recognition. A store might want a system to detect if someone is carrying a gun or stealing an item, for example.

Thus, although facial recognition technologies have animated much of the popular discourse around risks with emerging AI systems, HCCV more comprehensively encompasses a diverse array of computer vision models whose development raises difficult questions around privacy and fairness.

In addition, throughout this Article, I will use the terms "model" and "algorithm" largely interchangeably to refer to the machine learning model being used. "Algorithm" is technically a more expansive term, referring to a "set of rules a machine (and especially a computer) follows to achieve a particular goal."[35] "Machine learning model" ("model" for short) is a more precise term, but "algorithm" is more commonly used in colloquial discussions about AI (similarly, "machine learning" is a more precise term than "AI," but is less commonly used in colloquial settings). In general, I will use colloquial terms when referencing popular discourse. I will refer to "HCCV systems" to describe more expansively a particular product or service that includes an HCCV model.

Moving to the core terms for this Article, being "seen" refers specifically to having images of your face and/or body collected and processed for developing HCCV systems. This definition should encompass all computer vision contexts where there are privacy considerations under existing biometric information privacy laws, which will be further discussed in Section IV. Being "unseen" thus means *not* having your images collected or processed for developing HCCV. Note that being "seen"/"unseen" focuses specifically on the how the HCCV system is developed since the tension highlighted in this paper is between privacy and the desire to *develop* more accurate and fairer HCCV systems. The focus is *not* on the images collected during the deployment of the HCCV system or on whether HCCV systems should be deployed.

Being "mis-seen" refers to experiencing poor performance from a deployed HCCV system. This includes your face/body not being detected, being mis-recognized for someone else, someone else being mis-recognized for you, or having images/videos of you mis-classified or mis-characterized. The last category includes tasks like suspicious behavior detection, where you might be erroneously labeled as cheating on an exam or shoplifting. As will be explored in greater depth in Section V, the harms of being mis-seen are both absolute and relative. An HCCV system can be harmful because it performs poorly in certain scenarios for all people or because it performs more poorly for specific subgroups, potentially perpetuating stereotypes or creating discriminatory disparities.

Lastly, it is important to define the term "bias." Because "bias" is a catch-all term for many different types of disparities, some in the algorithmic fairness community have criticized the use of its term, arguing instead for more precise descriptions of the specific harms.[36] In this Article, I will use the term "bias" to refer to any case of disparate performance of the HCCV system across different groups, particularly when such disparate performance might lead to disproportionate harm for specific groups. When possible, however, I will refer instead to

[35] https://www.merriam-webster.com/dictionary/algorithm#note-1

[36] https://arxiv.org/pdf/2005.14050.pdf, https://arxiv.org/pdf/2103.06076.pdf (example delineating specific harms)

specific types of harms. The specific bias-related harms of being mis-seen are detailed in Section V. "Fairness" in this Article will refer to the pursuit of bias mitigation. It is impossible for an AI system to be completely unbiased or "fair," but the goal is to minimize bias as much as possible.

II: Why Computer Vision?

There is a general tension between privacy laws and algorithmic bias detection and mitigation efforts in that such efforts typically involve the use of protected class or sensitive attribute data, or proxies for such data. Prior works have discussed this empirically through interview methods[37] and in analyses of relevant antidiscrimination law prohibitions on the usage of such data.[38] This paper focuses on the context of bias mitigation in computer vision, given that here the concern is not simply with protected attributes or sensitive data, but rather with *all* of the data used in generating such models. In the tabular or language data contexts, stripping the dataset of personally identifiable information (PII) can significantly mitigate the privacy risks.[39] In contrast, for HCCV, even if the developer strips all of the metadata, the face or body images themselves constitute PII.[40] Moreover, developing HCCV generally involves the processing of biometric information, which is subject to strong privacy protections, as will be discussed further in Section IV.

Not only are the privacy concerns stronger in the HCCV context, but also the need for wide-ranging data collection efforts is greater. While a simple logistic regression model with tabular data can be trained on thousands of instances, HCCV requires millions of images to train a base model that can do basic detection and recognition tasks.[41] Moreover, while dataset diversity is important in all contexts, bias in computer vision is particularly strongly connected with a lack of sufficient dataset diversity. The primary technical solution to addressing the harms discussed below in Section V is to collect larger, more diverse, and more balanced datasets.[42] While there are other bias mitigation solutions that have been explored in the computer vision literature, these methods either rely on the generation of synthetic images to create a more diverse, balanced dataset[43] or address bias only indirectly.[44]

[37] https://arxiv.org/abs/2011.02282

[38] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3650635

[39] For tabular data, removing unique identifiers, employing differential privacy techniques, and limiting the number of and types of features are all techniques that can significantly reduce privacy concerns. Similarly, for language data, stripping the dataset of identifiers and contextual information, and limiting the amount of data from individual conversations can significantly reduce the ability to tie specific language data to individuals.

[40] https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-122.pdf

[41] Note, however, that it is fine for some subset of these images to be synthesized. https://arxiv.org/pdf/1603.07057.pdf

[42] https://arxiv.org/pdf/2004.07999.pdf; http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf; https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf; https://arxiv.org/pdf/2004.07999.pdf

[43] https://arxiv.org/pdf/2007.06570.pdf; https://arxiv.org/pdf/1810.00471.pdf; https://ieeexplore.ieee.org/document/7985521

[44] These types of methods work by focusing the model's attention on relevant features instead of irrelevant ones. See, e.g.,

In the tabular data context, the fundamental algorithmic bias problem is the presence of biased patterns in the data. For example, in criminal justice data, there is evidence that Black individuals have faced higher rates of arrest for drug-related crimes despite similar rates of offending.[45] Algorithms designed to predict recidivism thus can improperly learn to associate features correlated with being Black with higher rates of recidivism. The solution to such problems of biased historical data is not to gather more arrest data on Black individuals but rather to attempt to measure the impact biases have had on leading to higher arrest rates and counteract those biases in the data (e.g., through algorithmic rebalancing across groups[46] or trying to find less biased features for predicting criminal offense rather than arrest[47]). In contrast, the literature on bias in computer vision systems primarily points to the need for more diversity and representation of minority groups in the data used to develop such systems.[48] Algorithmic bias in the computer vision context generally boils down to two problems: lack of representation[49] and spurious correlations.[50]

The former refers to the lack of sufficient images of particular subgroups in a training dataset. This source of bias is also present in human face recognition, where studies have shown that people have a harder time recognizing people of other races.[51] Psychological research has also shown that the ability of humans to recognize faces of people of other races improves with more contact with people of other races.[52] Similar to humans, facial recognition algorithms also exhibit an “other-race effect,” whereby algorithms developed in Western countries perform better for Caucasian faces and algorithms developed in East Asian countries perform better for East Asian faces.[53] If you think of the machine learning developer as the parent to the HCCV system, a parent who wants to ensure their “child” is able to equally recognize people of all different races, it is easy to understand the urgency for collecting a diverse set of faces for training the algorithm.

The other fundamental source of bias is spurious correlations, meaning that the training data contain misleading patterns, often due to societal biases.[54] For example, researchers have found that gender classification models are more likely to incorrectly predict that an individual in a photo is female if the background is indoors and conversely for outdoor images.[55] Even though

https://openaccess.thecvf.com/content_ECCV_2018/papers/Lisa_Anne_Hendricks_Women_also_Snowboard_ECCV_2018_paper.pdf

[45] Sharad Goel, Justin M Rao, Ravi Shroff, et al. 2016. Precinct or prejudice? Understanding racial disparities in New York City’s stop-and-frisk policy. The Annals of Applied Statistics 10, 1 (2016), 365–394; Kristian Lum and William Isaac. 2016. To predict and serve? Significance 13, 5 (2016), 14–19;
Emma Pierson, Camelia Simoiu, Jan Overgoor, Sam Corbett-Davies, Daniel Jenson, Amy Shoemaker, Vignesh Ramachandran, Phoebe Barghouty, Cheryl Phillips, Ravi Shroff, et al. 2020. A large-scale analysis of racial disparities in police stops across the United States. Nature human behaviour (2020), 1–10.

[46] https://fairmlbook.org/pdf/fairmlbook.pdf

[47] https://arxiv.org/pdf/2105.04953.pdf

[48] https://arxiv.org/pdf/2004.07999.pdf; http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf; https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf

[49] http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf; https://arxiv.org/pdf/1906.02659.pdf

[50] https://arxiv.org/pdf/2004.07780.pdf; https://arxiv.org/pdf/1803.09797.pdf

[51] https://www.scientificamerican.com/article/some-people-suffer-from-face-blindness-for-other-races/

[52] Note, however, that this improvement only occurs up to the age of 12—greater social contact with people of other races in adulthood has little effect. https://www.nature.com/articles/s41598-019-49202-0

[53] https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=906254

[54] This is related to the problem of short-cut learning. https://arxiv.org/pdf/2004.07780.pdf

[55] https://arxiv.org/pdf/2004.07999.pdf

the background of an image should not be relevant for discerning whether an individual is male or female, models learn to rely on such irrelevant factors when the training data disproportionately features images of females indoors and males outdoors. Thus, it is important to develop training datasets that are well-balanced and avoid spurious correlations. For example, the proportion of women indoors vs. outdoors should roughly match the proportion of men indoors vs. outdoors. Of course, it is impossible to account for all spurious correlations, so researchers typically focus on ones that are related to pernicious societal stereotypes. Collecting a balanced dataset in an unbalanced world, however, can be difficult in practice, as the next Section will discuss.

In addition to bias mitigation, the prosocial normative motivation for collecting large, diverse datasets in computer vision is particularly strong given that doing so can directly improve the accuracy of the model.[56] Outside of the HCCV context, bias mitigation itself can be a source of controversy.[57] For example, in prior work, I have explored the ways in which many of the dominant approaches to bias mitigation in the tabular data ML context would be considered as legally suspect affirmative action.[58] In contrast, ensuring that your model recognizes people based on their facial features and not based on their clothing or the image background is important not only for reducing bias but also for increasing accuracy across a wider set of deployment contexts.[59]

That said, many of the insights from this Article are not unique to computer vision. If we can reconcile the tensions between privacy and antidiscrimination in HCCV, we might be able to apply analogous solutions to other forms of AI.

III: CHALLENGES TO ALGORITHMIC BIAS MITIGATION IN COMPUTER VISION

[56] Note that this is specifically true for verification and recognition tasks. For classification tasks, there can still be a trade-off between fairness and accuracy due to biases or stereotypes reflected in the classifications. https://pure.mpg.de/rest/items/item_3286836/component/file_3286837/content.

[57] https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/ (discussion about fairness-accuracy trade-off)

[58] https://lawreviewblog.uchicago.edu/2020/10/30/aa-ho-xiang/, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3650635

[59] The reason for this distinction is that in the tabular data context, one of the fundamental problems is a lack of ground truth. For example, there are significant concerns around the use of ML recidivism risk assessment tools in the criminal justice context given that arrest is used as a proxy for re-offense. https://arxiv.org/pdf/2105.04953.pdf Given the evidence that minority communities face disproportionately higher levels of policing and arrest, including wrongful arrest, arrest data itself is seen as suspect. The problem with having unreliable or biased arrest data, however, is that correcting for this bias requires some notion of what the data would look like in a fairer world. One extreme notion of this is the demographic parity fairness metric, which defines bias as disproportionate outcomes (in this case, predicted recidivism rates) across groups. Correcting for this bias metric would involve ensuring that the ML model predicts proportional recidivism rates across groups, even if the training data suggest highly disproportionate rates. Other approaches to bias mitigation take a more nuanced approach, but most are analogous to affirmative action in contemplating some degree of rebalancing across groups for fairness rather than accuracy purposes. In the computer vision context, however, ground truth is more readily accessible, so it is easier to align fairness and accuracy. For example, if the task is to verify whether two faces are of the same person, and the test set includes unique identifiers for each of the individuals, then correcting for problems of bias (e.g., the model being worse at distinguishing between individuals of darker skin tones) directly improves accuracy as well.

Collecting larger, more diverse training datasets and test datasets serves two aims: improving the overall accuracy and robustness of the model and also mitigating potential biases. While this Article addresses both of these aims, the focus is primarily on issues of bias since there are arguably sufficient existing commercial incentives to improve the overall performance of HCCV systems. Indeed, the accuracy of major commercial facial recognition technologies has improved dramatically over the past few years, but issues of bias persist.[60]

While the desire to build larger and more diverse datasets for training and testing computer vision systems is admirable, doing so immediately runs into complex questions of privacy, consent, money, and possible exploitation. Indeed, the computer vision community is infamous for blurring or crossing ethical lines in order to collect the large corpuses of data needed to train their systems. NIST uses images of mugshots, exploited children,[61] individuals crossing the border, and visa applicants in its test dataset, which is used to benchmark the performance of different commercial FRT.[62] Chinese start-ups have developed facial analysis systems for identifying ethnic minorities for surveillance purposes using “face-image databases for people with criminal records, mental illnesses, records of drug use, and those who petitioned the government over grievances.”[63]

Large publicly available human image datasets mostly use web-scraped photos. Some focus on celebrities or public figures (e.g., MS-Celeb-1M[64]); others focus on a broader array of subjects through online platforms like Flickr (e.g., YFCC100M[65]). Images of celebrities have especially assisted with the advancement of research into face recognition and verification systems since such datasets include many images of the same person, at different angles and in different settings. Such datasets, however, raise ethical issues around consent and also biases introduced by only training algorithms to recognize celebrities, whose features are not representative of the general population.[66]

[60] https://www.gao.gov/assets/gao-20-522.pdf

[61] These images are used specifically to test the performance of face detection and recognition systems on children. https://www.nist.gov/programs-projects/chexia-face-recognition. Images of children are hard to come by in most datasets due to additional privacy restrictions.

[62] https://slate.com/technology/2019/03/facial-recognition-nist-verification-testing-data-sets-children-immigrants-consent.html; https://pages.nist.gov/frvt/reports/11/frvt_11_report.pdf at 41-42; https://pages.nist.gov/frvt/reports/1N/frvt_1N_report.pdf (“The evaluation uses six datasets: frontal mugshots, profile view mugshots, desktop webcam photos, visa-like immigration application photos, immigration lane photos, and registered traveler kiosk photos.”)

[63] https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html

[64] https://www.microsoft.com/en-us/research/publication/ms-celeb-1m-dataset-benchmark-large-scale-face-recognition-2/ (featuring 10 million face images or nearly 100,000 individuals).

[65] http://projects.dfki.uni-kl.de/yfcc100m/ (featuring around 100 million images and videos).

[66] One artifact of using datasets exclusively of celebrities is that if you train a model to synthesize more feminine faces, it will do so by applying makeup to the face (specifically, a smokey eye and lipstick). https://dl.acm.org/doi/pdf/10.1145/3422841.3423533; https://arxiv.org/pdf/1812.00099.pdf. Looking more feminine is thus conflated with wearing makeup. In contrast, the models that synthesize more masculine features actually change the features of the face to be more angular. Datasets like CelebA that include an “attractiveness” feature are also problematic in that they can replicate human biases around what looks attractive. One study illustrated this by increasing the “attractiveness” latent attribute of Barack Obama, only to find that it made him look like a young, blonde white woman. https://arxiv.org/pdf/1907.12917.pdf

The use of Flickr images has been very pervasive in the computer vision community due to the uniquely diverse and candid nature of these images, which often include a wide variety of people and objects in each image. In fact, researchers who have constructed large public datasets using Flickr images have often been motivated to use Flickr to address the issues of bias that plague other datasets.[67] Flickr-based datasets features photos of non-celebrities[68] from amateur photographers,[69] yielding a large amount of diversity.[70] Recently, however, there have been many lawsuits leveraging Illinois' Biometric Information Privacy Act (BIPA) since the individuals in the Flickr images did not necessarily consent to having their photos used to train facial recognition algorithms.[71] Informed consent is thus a key consideration when collecting or using large image datasets for developing HCCV.

A. Why is Collecting Images with Informed Consent So Difficult?

The most obvious and reliable way to address the privacy concerns around collecting images for training HCCV systems is to obtain informed consent from the individuals in the photos. This is much easier said than done, however, given the need for millions of images with diverse subjects and conditions.

Social media or cloud service companies can collect large image datasets through products that incentivize individuals to upload photos. This is not to say they have always appropriately obtained informed consent, however. For example, Facebook recently reached a landmark settlement of $650 million in a BIPA case challenging their use of users' face images for training their face-tagging algorithm.[72] That said, for companies with a business model where individuals upload large numbers of diverse photos, the first step to solving the informed consent issue is comparatively straight-forward: Facebook now asks users whether they consent to having their images used for face tagging.[73]

This is not to say that the problem is completely solved—users upload many photos of people other than themselves. Even if the user has consented to the photos being used for facial recognition, providing that service still requires obtaining the consent of the individuals in the photo. Even if the individuals in the photo have Facebook accounts and have provided approval on their end, it is unclear how Facebook can know whether the individuals in the photo have given consent without first attempting to recognize the individuals.

[67] https://homes.cs.washington.edu/~kemelmi/ms.pdf; https://arxiv.org/pdf/1901.10436.pdf; https://arxiv.org/pdf/1405.0312.pdf

[68] https://homes.cs.washington.edu/~kemelmi/ms.pdf

[69] https://arxiv.org/pdf/1405.0312.pdf

[70] https://arxiv.org/pdf/1901.10436.pdf

[71] https://www.nbcnews.com/tech/internet/facial-recognition-s-dirty-little-secret-millions-online-photos-scraped-n981921; https://www.reuters.com/article/dataprivacy-ibm-biometrics/ibm-cant-shake-facial-recognition-suit-but-dodges-some-claims-idUSL1N2GD2JP; https://techcrunch.com/2020/07/15/facial-recognition-lawsuit-vance-janecyk-bipa/

[72] https://apnews.com/article/technology-business-san-francisco-chicago-lawsuits-af6b42212e43be1b63b5c290eb5bfd85; https://www.law.com/therecorder/2021/02/26/judge-approves-facebooks-650m-privacy-settlement-as-major-win-for-consumers/

[73] https://www.facebook.com/help/122175507864081

For companies without business models that incentivize data collection organically, the need to collect large, diverse datasets with informed consent poses additional difficulties. The companies can buy images from vendors, but it is difficult to (i) obtain enough data and (ii) obtain sufficiently diverse and candid data. While Facebook might not have to pay users to upload thousands of pictures of themselves and their friends, a company using a vendor to collect images must pay for each image. To collect the millions of images needed to train from scratch a computer vision model with good performance, millions of dollars is generally needed.

In addition, when crowdsourced workers are paid to upload images of themselves based on particular specifications (e.g., one front-facing photo, one side-facing photo, one photo indoors, one photo outdoors), the photos generally look staged.[74] In computer vision, there is a term "in the wild," which refers to "unconstrained" images that appear to be taken in a wide variety of everyday scenarios—similar to the contexts a deployed HCCV system would be working with.[75] When buying photos, however, it can be difficult to find appropriate photos "in the wild." It is much harder to ensure such photos meet particular specifications, verify that all the individuals in the photos have consented to the use of the photo for training AI, and verify that the photographer has relinquished their copyright.

These challenges create a number of performance and bias concerns. First, an HCCV model trained on very staged selfies might struggle to perform in the real world, where there might be multiple people in an image, the lighting conditions might be more varied, the people might be smaller and blurrier, or the people might have a wider variety of poses or expressions. Moreover, if the dataset features images from only one country— often the case given the need for the crowd workers to sign a consent form based on the laws of their jurisdiction—then there might be issues of bias in the dataset. Not only might there be insufficient demographic diversity, but also the backgrounds and objects in the photos might only reflect country-specific contexts.[76]

In addition, it can be more difficult to obtain as much metadata about the images as might be available to platforms with users uploading photos. For example, Facebook has a large amount of information about individuals beyond simply their faces—Facebook might know where a photo was taken[77] and the demographics and interests of the individuals in the photos.[78] This leads to a richer dataset that can be used for profitable tasks like ad targeting.[79]

I emphasize this distinction between the challenges faced by companies with platforms where people upload images freely versus other companies because this creates competition concerns in addition to the privacy and bias concerns discussed elsewhere. There are relatively few companies that have the advantage of a large, global, diverse usership willing to upload

[74] In the early days of developing computer vision datasets, researchers did stage the photos they collected, hiring actors and photographers, and manually designing the set-up. https://arxiv.org/pdf/2102.00813.pdf This was a very labor-intensive and expensive process, so early datasets were quite small. The need for informed consent, however, raises the question of how we can adapt these more manual ways of collecting images to suit the needs of contemporary computer vision development.

[75] http://vis-www.cs.umass.edu/papers/lfw.pdf

[76] https://research.fb.com/wp-content/uploads/2019/06/Does-Object-Recognition-Work-for-Everyone.pdf

[77] https://www.facebook.com/help/115298751894487/. A company working with a vendor can also access location data if the EXIF metadata on the photo has not been stripped. It is, however, common to strip such data (which includes timestamps and GPS coordinates) since it can create additional privacy concerns.

[78] https://www.facebook.com/business/ads/ad-targeting

[79] *Id.*

millions if not billions of photos for free. There are far more companies that either operate or seek to operate in the HCCV space.

Moving beyond the necessity to collect large numbers of images, the need to collect a diverse, well-balanced dataset with minimal spurious correlations creates additional challenges. First, there is the challenge of defining what sufficient diversity would look like. Relevant dimensions of diversity from the computer vision literature include demographics (perceived gender, age, and ethnicity), lighting conditions, background, pose, and camera type.[80] Avoiding spurious correlations would mean ensuring that no unrelated attributes are inadvertently correlated—e.g., women and men in equal proportion indoors vs. outdoors. In addition, determining the relevant subcategories within each group is a challenging task that AI developers are not necessarily the best equipped to determine. For example, should there be two gender categories? Three? Ten?

Even after these sociological questions are answered about the ideal taxonomy and distribution for the dataset, there is the challenge of fulfilling these specifications. When issues of bias are discovered in the dataset or in models trained on it, it can be difficult to augment the dataset to address these issues. For example, if a developer realizes that their model does not perform well for Native American individuals, and they check their training dataset and realize they do not have any images of Native Americans, a natural solution would be to seek out images of Native Americans. Conducting that type of targeted recruitment can be very difficult. Especially when collecting data from historically marginalized communities, it is important to ensure that the data collection process is not exploitative.

Moreover, publicly available datasets typically do not include people's self-reported demographics, so researchers or developers take measures to guess or estimate the demographics. Datasets with celebrities sometimes have web-scraped data on nationality.[81] When that information is not available, common methods include having annotators look at the photos and guess people's demographics,[82] using skin tone or other features as a proxy for race,[83] or using automated race classifiers.[84] While it would be much more ideal to collect the self-reported demographics of the image subjects, collecting demographic data can present additional privacy concerns, as will be discussed further in Section IV. Without such data, however, even doing a preliminary check to see if the dataset is diverse or if the model performs well across different groups is difficult.[85]

Given all of these data collection challenges, even the computer vision research community is divided on how important informed consent should be for image datasets.[86] More than half of the respondents to a survey conducted by Nature did not think it was necessary to obtain informed consent from individuals before using their face images. Even researchers who

[80] https://arxiv.org/pdf/1810.03993.pdf

[81] https://arxiv.org/pdf/1812.00194.pdf; https://dl.acm.org/doi/pdf/10.1145/3422841.3423533

[82] https://arxiv.org/pdf/1908.04913.pdf;

[83] https://arxiv.org/pdf/1901.10436.pdf

[84] https://arxiv.org/pdf/1812.00194.pdf

[85] Without such data, companies often rely on proxy variables. Andrus et al. For example, skin tone might be used as a proxy for race, or long hair as a proxy for gender. There are many downsides, however, to using such proxies. See Gender Shades (discussing shortcomings of using the Fitzpatrick skin tone scale as a proxy for race); Xiang (discussing unintended consequences of using proxy variables for bias mitigation). https://arxiv.org/pdf/1812.00099.pdf (differences in performance are not due to skin tone)

[86] https://www.nature.com/articles/d41586-020-03187-3

believed in the importance of informed consent stated they would still use datasets that do not have appropriate informed consent. It was difficult for the researchers to see how they could train accurate facial recognition algorithms otherwise.

Overall, the challenge of assembling large, diverse, and well-balanced human image datasets is a topic that requires more public awareness. When an AI system fails to work well for individuals from marginalized backgrounds, this often becomes a source of public outrage and used as evidence that companies do not care about such individuals. Even in situations where people do care deeply about making their products work well for everyone, however, collecting sufficiently large and diverse datasets is very difficult and runs directly into many privacy challenges.

IV: Privacy Laws

There are two separate areas of privacy law that are relevant to the context of mitigating bias in computer vision systems: (i) the collection of PII (in particular, biometric information) and (ii) the collection of sensitive attributes. The former is generally relevant for the development of any HCCV system, but raises particular concerns in the context of attempting to collect more diverse datasets, focusing on marginalized groups. The latter is important for both bias detection and mitigation; you cannot evaluate dataset diversity or performance across demographic groups without demographic information.

Some of the most salient privacy laws in the first category are GDPR's restrictions around the processing of personally identifiable information (PII)[87] and U.S. state laws like BIPA[88] and the California Consumer Privacy Act (CCPA)[89] that regulate the processing of biometric information.[90] Biometric information in this context can be seen as a particularly sensitive subset of PII. BIPA, for example, regulates the collection, storage, and use of biometric identifiers and biometric information. Biometric identifiers include "scan[s] of hand or face geometry," which has been interpreted by courts to include both facial landmarks and facial templates, which are extracted for any computer vision task involving detecting, verifying, recognizing, or classifying faces. CCPA's protections of biometric information include face images, images of hands or palms, and gait patterns.

While each state's biometric information privacy laws differ slightly in scope, they generally all seek to restrict the collection, storage, and use of images/videos of faces or bodies (or landmarks/templates extracted from these images/videos) that *could* in turn be used to identify a person (actual use for identification is not required). The laws vary in terms of the rights they provide; some provide a right to request and receive disclosures about information that has been collected (CCPA, BIPA), a right to request that the information be deleted (CCPA), a prohibition of denying goods or services for exercising privacy rights (CCPA), or a prohibition

[87] https://gdpr.eu/eu-gdpr-personal-data/

[88] https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57

[89] https://oag.ca.gov/privacy/ccpa

[90] Texas and Washington have also passed biometric information privacy laws. https://statutes.capitol.texas.gov/Docs/BC/htm/BC.503.htm; https://app.leg.wa.gov/RCW/default.aspx?cite=19.375.

on sale of or profit from the information (BIPA).[91] The key protection this Article will focus on, however, is the requirement of informed consent in order to collect biometric information. This is the one constant across the various laws, and, as discussed above in Section III, creates significant challenges in the development of HCCV.

Most of the U.S. privacy cases about image collection for HCCV center on BIPA. BIPA was passed in 2008, making it much older than other comparable state laws. Over the past several years, BIPA's private right of action has made it a powerful tool for privacy advocates to challenge tech company data practices. Although state statutes like BIPA in theory are narrow in their jurisdictional scope, in practice the difficulty of determining whether images in a dataset are from Illinois residents vs. other states' residents has vastly expanded the influence of BIPA.[92] Section V will feature a more in-depth discussion about the specific harms these laws seek to prevent and how courts have interpreted them.

GDPR adds an additional layer of complexity by requiring not only consent but also a legal basis for processing personal data, which includes biometric information. For example, Sweden's first GDPR fine was issued against a school that used facial recognition to take attendance.[93] Although parents had consented for the use of such technology for their children, that did not settle the legal question of necessity. The Swedish Data Protection Authority concluded that there were less intrusive ways to take attendance than camera surveillance given the expectation of privacy students would have in the classroom.

In the second category—laws protecting sensitive attribute data—we again have GDPR, which regulates the processing of special categories of personal data like race.[94] We also have some U.S. privacy laws and antidiscrimination laws, like the Equal Credit Opportunity Act, which place additional restrictions on the collection or consideration of sensitive demographic data.[95]

In practice, these restrictions have ironically erected significant barriers to companies attempting to self-audit their algorithmic systems for bias. In a recent interview study,[96] my co-authors and I found that overwhelmingly companies across the AI industry, both small and large, struggle to check their AI systems for bias, let alone take remedial measures to address bias. Despite the growth in AI ethics, responsible AI, and algorithmic fairness teams in tech companies, these teams face practical challenges when attempting to convince their colleagues to collect sensitive attribute data in order to conduct bias assessments. Often legal and compliance teams shut down efforts to collect, share, or use such data. In light of existing privacy laws, this knee-jerk reaction is understandable, but in practice, it makes progress toward less-biased AI more challenging.

[91] https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57; https://oag.ca.gov/privacy/ccpa

[92] Some companies have tried to sidestep BIPA and other state information privacy laws by asking individuals what state they are residents of before giving them access to a product. https://newmedialaw.proskauer.com/2018/01/18/google-app-disables-art-selfie-biometric-comparison-tool-in-illinois-and-texas/. Note, however, that Google did ask for consent from users of the app before processing their selfies

[93] https://www.bbc.com/news/technology-49489154

[94] Article 9, GDPR, https://gdpr-info.eu/art-9-gdpr/

[95] Regulation B, § 1002.5(b), https://www.consumerfinance.gov/rules-policy/regulations/1002/5/#a-4-vi

[96] *See* Andrus et al. supra note []. https://dl.acm.org/doi/pdf/10.1145/3375627.3375852 also discusses the challenge in practice of balancing data minimization under GDPR and bias audits.

There is evidence that policymakers are increasingly cognizant of this challenge. The E.U. proposed AI regulation creates a carve-out for processing sensitive data for bias monitoring, detection, and correction for high-risk AI systems.[97] In addition, the UK's Information Commissioner's Office (ICO) has released guidance suggesting that such data can and should be collected for the purposes of bias mitigation, and recommends pursuing the public good exception in GDPR.[98] Less progress has been made on the U.S. side, however. While there have been growing calls for audits of tech company algorithms,[99] there has been less policy discussion around ways to better *enable* companies to conduct audits. More generally, there seems to be less recognition of the existence of this tension between existing U.S. privacy and antidiscrimination laws and the pushes for less-biased facial recognition systems.[100]

In short, while there is growing recognition that sensitive attribute data might be needed for bias detection and mitigation, collecting large and diverse datasets that comply with privacy laws remains a major challenge.

V: Harms of Being Seen vs. Mis-seen

One of the core contributions of this Article is to identify and characterize the tension between protecting against the harm of being "seen" by HCCV systems versus the harm of being "mis-seen" by such systems. The former is the primary concern of privacy law, whereas the latter is the primary concern of the algorithmic fairness community. Since both are important ethical considerations, this section will focus on breaking down the specific harms of being "seen" and "mis-seen" in order to better delineate the potential trade-offs involved.

A. Harms of Being Seen

Privacy law is notorious for the ambiguity around the specific harms it envisions. In the seminal article "The Right to Privacy," which is credited for essentially creating the U.S. common law privacy right,[101] Warren and Brandeis discuss privacy as "the right to be let alone."[102] The authors compare privacy to "the right not to be assaulted or beaten, the right not to be imprisoned, the right not to be maliciously prosecuted, the right not to be imprisoned, the

[97] "To the extent that it is strictly necessary for the purposes of ensuring bias monitoring, detection and correction in relation to the high-risk AI systems, the providers of such systems may process special categories of personal data . . ." Title III, Chapter 2, Article 10.5, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

[98] https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-ai-and-data-protection/what-do-we-need-to-do-to-ensure-lawfulness-fairness-and-transparency-in-ai-systems/

[99] https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1212&context=penn_law_review_online; https://hbr.org/2018/11/why-we-need-to-audit-algorithms; https://www.brookings.edu/research/auditing-employment-algorithms-for-discrimination/

[100] *See* Andrus et al., supra note [].

[101] Daniel J. Solove & Paul M. Schwartz, Information Privacy Law 10-11 (7th ed.).

[102] Samel D. Warren & Louis D. Brandeis, The Right to Privacy, 4 Harv. L. Rev. 193 (1890).

right not to be defamed."[103] In contrast to the laws governing those rights, the authors conceived of privacy as protecting against mental suffering rather than simply reputational damage (as under defamation law) or infringements upon property (as under intellectual property law).[104] They justify privacy protections as an extension of common law's "secur[ing] to each individual the right of determining to what extent his thoughts, sentiments, and emotions shall be communicated to others."[105]

Modern consumer data privacy law is rooted in tort law, contract law (when companies employ privacy policies), property law, Section 5 of the FTC Act (prohibiting "unfair or deceptive acts or practices in or affecting commerce"), sectoral federal statutory regulation, and state statutory regulation.[106] Most relevant to our discussion, however, are state biometric privacy laws like Illinois' BIPA and California's CCPA. These laws go beyond the sectoral nature of federal privacy laws and provide protections for biometric information or personal data, regardless of the context of collection or use. While the right to privacy writ large might be conceived of as a right to be left alone, biometric privacy laws specifically protect an individual's control over their data, making informed consent the key requirement for collection, storage, or use.[107]

What, however, are the specific harms that laws like BIPA protect against? In *Spokeo v. Robins*, the Court found that mere violation of BIPA alone did not constitute a concrete injury sufficient for standing; the requirements of standing had to be independently met.[108] In a subsequent case, *Patel v. Facebook*, the Ninth Circuit developed a test for whether a statutory violation caused a concrete injury: "(1) whether the statutory provisions at issue were established to protect concrete interests (as opposed to purely procedural rights) and, if so, (2) whether the specific procedural violations alleged in this case actually harm, or present a material risk of harm to, such interests."[109] Applying this test to the context of Facebook using facial recognition in its "Tag Suggestions" technology, the Ninth Circuit determined that, "the development of a face template using facial-recognition technology without consent (as alleged here) invades an individual's private affairs and concrete interests." The court thus found that BIPA protects an individual's concrete privacy interests, such that violations of the procedures in BIPA "actually harm or pose a material risk of harm to those privacy interests."

Key to the Ninth Circuit's decision was the idea that common law protects an individual's "control of information concerning his or her person," such that lack of control over one's biometric information, as protected against by BIPA, constituted a concrete harm. Lack of control over data is still quite a broadly construed harm, however. In the HCCV cases litigated thus far in the U.S., there has been no allegation of a disclosure of information leading to mental harm similar to the gossiping press that was decried by Warren and Brandeis as the impetus for a

[103] *Id.*
[104] *Id.*
[105] *Id.*
[106] Solove & Schwartz, supra note [] at 812-813.
[107] BIPA, for example, prohibits private entities from collecting, capturing, purchasing, receiving through trade, or otherwise obtaining a person's or a customer's biometric identifier or biometric information without first (i) informing the individual that the biometric identifier or information is being collected or stored, (ii) informing the individual of the length of time of the collection, storage, or use, and (iii) receiving written release from the individual.
[108] Spokeo v. Robins, 578 US _ (2016).
[109] Patel v. Facebook Inc., 290 F. Supp. 3d 948 (N.D. Cal. 2018).

right to privacy. Instead, in evaluating Article III standing, U.S. courts have found compelling the concrete harms of identity theft and surveillance risk.[110]

The concern around identity theft is that as face verification is increasingly used for security purposes (e.g., opening phones, accessing buildings, payment), face templates extracted from images could be used to gain unauthorized access. For example, in *Patel v. Facebook*, the court expressed concern that the face templates collected by Facebook could be used to unlock cell phones.[111] It is unclear, however, that extracting face templates or landmarks from face images in order to develop HCCV increases the security risks beyond simply storing the images themselves. Existing methods to hack face verification systems rely on generating 3D renderings using publicly available images of the individual being hacked.[112] Especially if the developer is using publicly available images to develop the HCCV system, it is unclear that doing so would increase the risk of identity theft for the image subjects, making this harm quite speculative.

The more significant potential harm animating privacy fears around HCCV is the specter of mass surveillance. Despite the fact that mass surveillance is characterized as a core privacy harm, however, it actually encompasses both harms of being "seen" and being "mis-seen."

Starting first with the mass surveillance harms of being "seen," it is helpful to distinguish between the use of HCCV for relatively low-stakes tasks like ad targeting versus higher-stakes purposes like law enforcement, employment, and finance. In low-stakes contexts, mass surveillance is primarily perceived by consumers as "creepy."[113] Calls for greater privacy protections are typically motivated by concerns that large tech companies are learning too much about individuals' personal lives. In a commonly cited example, Target sent targeted ads featuring maternity and baby products to a household with a teenage daughter.[114] The father was outraged that Target would send such materials to his daughter and complained to the company, only to later retract his complaint after he realized his daughter was indeed pregnant.

While concerns about technology companies being "creepy" and knowing too much about us are highly relatable and concerning insofar as they lead to lack of freedom of expression or self-censorship,[115] discussions around the harms of mass surveillance from FRT[116] tend to

[110] In a recent lawsuit against TikTok, the court also considered the economic harm of use of electricity and processing power. This is relevant to cases where companies train their HCCV systems in the background of the plaintiff's phones, computers, and other devices. This harm, however, is not characteristic of all HCCV systems—it depends on whether the company is using the edge device (e.g., the individual's phone, computer, or camera) for training vs. collecting the images and then training the model on their own servers. https://www.intel.com/content/www/us/en/edge-computing/edge-devices.html

[111] https://cases.justia.com/federal/appellate-courts/ca9/18-15982/18-15982-2019-08-08.pdf?ts=1565283704

[112] https://www.wired.com/2016/08/hackers-trick-facial-recognition-logins-photos-facebook-thanks-zuck/

[113] https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.851.3914; https://dl.acm.org/doi/pdf/10.1145/3173574.3174067. Notably, even in the ad targeting contexts, one of the principal harms is also of being mis-seen. For example, in an interview study, users complained about being stereotyped by online behavioral advertising. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.851.3914

[114] https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

[115] See, e.g., http://www.bu.edu/jostl/files/2020/08/1-Barrett.pdf. This is related to the idea of privacy as a personality right, as captured in the original Warren and Brandeis article. They characterized the freedom afforded by privacy as allowing individuals "to give full expression to his peculiar capacities and powers." Note, however, that this is not officially a right associated with privacy in the U.S., in contrast to Germany. Solove & Schwartz, supra note [] at 24-25.

[116] I use the term FRT rather than HCCV in this discussion since the discourse I am referring to specifically relates to FRT.

center instead on concerns specific to the sharing of information with law enforcement.[117] Being "seen" by law enforcement can be good for society from the perspective of assisting police in capturing the perpetrators of crimes. Indeed, much of the incentive for the proliferation of both private surveillance systems and law enforcement FRT use has been the utility of such technologies in solving crimes and catching culprits.[118]

Much of the criticism of law enforcement use of FRT thus centers not on the harms of being "seen" but rather the harms of being "mis-seen." While FRT are constantly improving in accuracy,[119] the potential for higher rates of mis-recognition of women and minorities remains. Although humans also do not have perfect facial recognition accuracy—indeed, eyewitness testimony can be highly flawed and manipulable[120]—critics of FRT have argued that there are currently insufficient safeguards in place to ensure that the technology is used appropriately and does not further compound existing trends of over-policing of minority communities.

This is not to say that being "seen" is not also considered a relevant harm. In contexts where there is significant distrust of the government or disagreement about the appropriateness of the laws being enforced, being "seen" is also considered a societal harm. For example, there has been significant criticism of government efforts to surveil journalists[121] or opposition party members.[122] Moreover, one of the most controversial uses of mass surveillance is the Chinese government's tracking of Uyghur minorities.[123]

Overall, however, it is important to recognize that while the harms of mass surveillance might seem at first blush to center on the harms of being "seen," in practice the core concerns cited in policy discussions are often harms of being "mis-seen." As this Article highlights, while privacy law protects individuals against being "seen," it can ironically make it more difficult for individuals to avoid being "mis-seen."

Moreover, while preventing mass surveillance has been a major motivation for strict privacy laws around processing biometric information, not all forms of HCCV necessarily facilitate mass surveillance. Face/body/object detection do not directly enable mass surveillance since they do not involve identifying individuals. Moreover, whether recognition technologies enable mass surveillance depends on the degree to which the data on face/body matches is shared. If FRT is used only locally on your phone to sort your photos, and the matches are not

[117]

[118] Studies have shown that participants tend to view the use of FRT as a trade-off between privacy and security. https://arxiv.org/pdf/2008.07275.pdf Of course, the actual utility of FRT for such purposes is highly disputable. See, e.g., https://www.smartcitiesdive.com/news/report-new-york-facial-recognition-pilot-flops/552399/; https://48ba3m4eh2bf2sksp43rq8kk-wpengine.netdna-ssl.com/wp-content/uploads/2019/07/London-Met-Police-Trial-of-Facial-Recognition-Tech-Report.pdf; https://www.nytimes.com/2020/01/12/technology/facial-recognition-police.html

[119] https://pages.nist.gov/frvt/reports/1N/frvt_1N_report.pdf

[120] https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=1075&context=psychology_pubs

[121] https://www.rcfp.org/nsa-mass-surveillance-against-journalist/; https://www.dni.gov/files/documents/RG/Effect%20of%20mass%20surveillance%20on%20journalism.pdf

[122] https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/download/6614/6466/; https://www.tde-journal.org/index.php/aseas/article/download/2648/2260

[123] https://www.washingtonpost.com/technology/2020/12/08/huawei-tested-ai-software-that-could-recognize-uighur-minorities-alert-police-report-says/; https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html

shared with the company or anyone else, then such technology arguably does not enable surveillance.

The goal of this Section is not to minimize the harms that privacy law seeks to protect against. Instead, the goal is to clarify that many of the concerns motivating the use of privacy law as a way to reign in FRT are rooted in fears of being "mis-seen" rather than being "seen."

B. Harms of Being Mis-seen

In this section, I will focus on four specific harms of being "mis-seen": differences in service provision, security threats, allocative harms, and representational harms. All of these harms are caused by differences in the performance of the algorithmic system for different groups (e.g., lower accuracy rates or higher false positives/negatives for women and minorities), but they are distinguished by how this difference in performance affects the individuals.

First, differences in service provision refer to contexts where an algorithmic system performs a function less well for certain groups versus others. For example, if a facial verification system is used at border control to determine whether an individual's face matches the photo in their passport, but that system is less accurate for Middle Eastern individuals, then Middle Eastern individuals are more likely to be flagged and sent to a separate line for a human to conduct the verification.[124] This is the most common and generic type of harm, and applies to virtually all computer vision tasks. In the face/body detection context, if an AI-assisted AC system is less proficient at detecting individuals with darker skin tones, then those individuals might find that the AC often turns off even when they are still in the room.

A second category of harm is security threats. This type of harm is specific to the verification context. For example, if the face verification algorithm on your phone is not very good at distinguishing between different Asian people, and you are Asian, then other Asian people might be able to unlock your phone. This is particularly a concern in households, where, bias aside, family members can sometimes unlock each other's phones.[125] Increasingly, face verification is also used for building security and for payments,[126] so significant discrepancies in the ability of such systems to work for different groups could lead to substantial security risks (e.g., someone breaking into your home or using your credit card).

The third category of harms is allocative harms. This is when an inaccuracy leads to a misallocation of a good or opportunity. In the computer vision context, this is most relevant to recognition and classification tasks. The example of wrongful arrest due to a faulty facial recognition match is a very high-stakes example of allocative harm, as individuals are unjustly deprived of their liberty. In terms of classification tasks, algorithmic systems that seek to identify suspicious behavior or categorize an individual's mental state or ability can also lead to significant allocative harm. For example, a study found that eye tracking devices did not work as

[124] Given harmful stereotypes about Middle Eastern individuals in the airport security context post-911, such service provision harms could lead to allocative harms if the individual is falsely accused of carrying a passport not belonging to them.

[125] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3620883

[126] See, e.g., https://ny.curbed.com/2019/10/8/20903468/nyc-facial-recognition-technology-homes-businesses; https://www.latimes.com/business/technology/story/2020-08-14/facial-recognition-payment-technology.

well for Asian participants as for other groups.[127] As such technology is increasingly used by educational institutions to determine whether students are paying attention and to detect cheating behavior,[128] such disparities in performance could lead to a higher risk of Asian students being incorrectly flagged for bad behavior.

Finally, we have representational harms, when algorithmic systems represent certain groups in negative, offensive, or other problematic ways. This type of harm is most relevant for classification tasks since such tasks involve applying a label to an image. A famous computer vision example of a representational harm was when Google Photos labelled an image of two Black individuals as an image of gorillas.[129] This harm can also occur with algorithms that determine which parts of images are the most relevant to focus on. Earlier this year, Twitter scrapped its image cropping algorithm following revelations that their algorithm was more likely to crop out black faces in favor of white faces.[130] Representational harms can also stem from existing biased trends in society. In the popular COCO dataset, images of women playing sports are more likely to be indoors, whereas images of men playing sports are more likely to be outdoors.[131] This can lead to HCCV models trained on COCO learning stereotyped representations. AI-powered image caption generators might consistently incorrectly label images of women playing outdoor sports as men and vice versa for men playing indoor sports, further perpetuating existing stereotypes.

While this Article primarily focuses on non-generative models, it is worth noting that representational harms are an especially relevant type of harm to consider when evaluating generative models. For example, Generative Adversarial Networks (GANs) trained to generate a synthetic image of an individual with longer hair have been shown to also feminize the facial features of the individual.[132] By conflating long hair with feminine facial features, the GAN perpetuates the stereotype that men have short hair and women long hair. Generative language models have also been shown to be vulnerable to generating highly racist and offensive language. For example, Microsoft famously scrapped its chatbot Tay after the bot started making highly inflammatory statements.[133]

Most concerns about bias in computer vision apply primarily to contexts where images of humans are used, but bias can also manifest itself in object detection or recognition. For example, researchers at Facebook found that their tool had a harder time identifying objects in photos taken in developing countries.[134] Because their training data was disproportionately collected from developed countries, the model could only recognize toothpaste on a sink in a more affluent-looking bathroom. This is why, depending on the task, it is important not only to

[127] https://www.researchgate.net/publication/236266469_Eye-tracking_data_quality_as_affected_by_ethnicity_and_experimental_design

[128] https://www.vice.com/en/article/n7wxvd/students-are-rebelling-against-eye-tracking-exam-surveillance-tools

[129] Notably this highly offensive harm seems to still not have been directly solved for. https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/

[130] https://www.cbsnews.com/news/twitter-kills-its-automatic-cropping-feature-after-complaints/; https://blog.twitter.com/engineering/en_us/topics/insights/2021/sharing-learnings-about-our-image-cropping-algorithm.

[131] https://arxiv.org/pdf/2004.07999.pdf

[132] https://authors.library.caltech.edu/107574/1/2007.06570.pdf

[133] https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

[134] https://research.fb.com/wp-content/uploads/2019/06/Does-Object-Recognition-Work-for-Everyone.pdf

consider the demographic diversity of the people in the images, but also to consider factors like the geographic diversity of where the images are taken.

1. Lack of Protections Against Being Mis-Seen

While being "seen" by an HCCV system without informed consent is considered under privacy law to be a harm in and of itself, being "mis-seen" is only considered to be a harm if it leads to a separate legally cognizable harm. For example, when Robert Williams sued the Detroit police department, he brought his action under the Fourth Amendment right to be free of unlawful seizures and the Elliot-Larsen Civil Rights Act, which protects against government entities "deny[ing] an individual full and equal enjoyment of public services on the basis of race.[135]

One could argue that this distinction is reasonable—that the harms of being mis-seen are already appropriately accounted for through existing anti-discrimination laws and other laws. Anti-discrimination law, however, primarily applies in specific, comparatively high-stakes contexts, like employment,[136] housing,[137] and finance.[138] While the limited domains of antidiscrimination law might be reasonable in the context of human discrimination, algorithmic discrimination raises additional concerns. The growing proliferation of HCCV in everyday life suggests that even small or subtle biases might accumulate into substantial harms.

Imagine, for example, being an individual of a minority demographic living in a world of HCCV designed for individuals in the majority group. Upon waking up, you check your phone, but it does not recognize you, so you have to manually input your passcode. Taking public transit to work, you try to use the facial recognition system to pay your fare, but it does not recognize you, so you have to go through a special line with a human verifier and arrive late to work. You join your colleagues for coffee at a cafe, but again the payment system fails to recognize you. You are embarrassed as the automated system says your face does not match the bank account you are trying to access, and you have to ask the cafe staff to give you another method of payment. They unfortunately do not have any other methods of payment, so you need to ask a colleague to cover your tab. When you and your colleagues return to the office, you are unable to enter the building, because the security system does not recognize you as one of the employees. While your colleagues are waiting for you, you call for a security guard to help you enter the building. The security guard is suspicious of your claim that you work in the office—the picture in the employee database looks like it *could* be someone else, and the AI system works extremely well for everyone else. Fortunately your colleagues vouch for you, and the security guard lets you in. At the end of the work day, you stay late, after your colleagues have left, to finish a project. The lights and AC turn off, as the AI-enabled AC and lighting systems do not detect any people in the office. Sitting in the darkness, you are confronted with your own invisibility.

[135] https://www.aclumich.org/en/press-releases/farmington-hills-father-sues-detroit-police-department-wrongful-arrest-based-faulty

[136] Title VII

[137] FHA

[138] ECOA

In the above scenario, I have only discussed a few of the possibly many instances of inconvenience, indignation, or embarrassment that might occur over the course of the day due to being "mis-seen" by HCCV. While most of the harms described would not be legally cognizable, together they amount to being treated as a second-class citizen, living in a world that cannot detect or recognize you. The sensation is similar to being a foreign tourist, forced to use alternative systems since you do not have a phone number, address, bank account card, etc. in the country, except that you cannot prevent these harms by simply setting up relevant accounts—you would need to change your face.

Of course, the scenario I described is extreme in that it is unlikely that most commercial AI systems would perform *so* consistently poorly for individuals in minority groups—occasional poor performance is much more likely. Nonetheless, currently the primary forces preventing such poor performance are the competitiveness of the market and the desire of companies to produce high-performing products. There is no legal protection for the individual in this scenario.

VI: Approaches to Balancing Privacy and Bias Mitigation

While privacy laws protect generally against the harms of being "seen" without consent, the harms of being "mis-seen" are not directly protected against. For practitioners charged with balancing the ethical desiderata of fairness and privacy, the threat of legal liability leans far more heavily in favor of protecting privacy than addressing algorithmic bias.[139] There are a few possible approaches for addressing this imbalance, as the subsections below will discuss.

One would be to decrease the protections against being "seen" through privacy law carve-outs. Another path would be to alleviate some of the concerns with being "seen" through participatory design, use of trusted third-parties to collect data, or privacy-preserving technological advances. Finally, a third approach would be to increase the protections against being "mis-seen."

A. Carve-Outs from Privacy Law

The idea of reducing privacy protections around FRT and other HCCV technologies might seem absurd at a time when there are calls for *stronger* privacy protections and the specter of mass surveillance seems increasingly threatening, with more and more deployment of HCCV technologies.[140] Indeed, some scholars have argued that we should instead be increasing privacy protections in the U.S. in order to prevent the ethical and legal risks associated with FRT.[141] Even China recently increased privacy protections in the FRT context. The Supreme People's

[139] See Andrus et al., supra note [].

[140] http://www.bu.edu/jostl/files/2020/08/1-Barrett.pdf, http://www.bu.edu/jostl/files/2017/04/Greenbaum-Online.pdf, https://repository.law.miami.edu/cgi/viewcontent.cgi?article=1345&context=umlr

[141] http://www.bu.edu/jostl/files/2017/04/Greenbaum-Online.pdf

Court issued a directive to lower courts to make "collection and analysis of facial data by companies an infringement of personal rights and interests if carried out without previous consent."[142]

Nonetheless, given the tension presented in this Article between multiple ethical desiderata—privacy and fairness—it is worth considering what possible surgical changes could be made to existing privacy regimes to balance these desiderata. Some countries, like Japan, have taken the approach of significantly weakening legal protections against data mining.[143] While such regulatory measures have been motivated by the desire to facilitate technological innovation, narrower carve-outs would likely better thread the needle of simultaneously maximizing privacy protections and minimizing the potential of being mis-seen by HCCV systems.

One possible carve-out is to make a distinction between images used to *develop* HCCV models versus images used during the *deployment* of an HCCV system. Training datasets are simply used to train the model to perform a specific task like detection, recognition, or classification in a particular context. While the HCCV system learns *how* to identify the individual, the goal is not to actually identify that individual. Similarly test datasets are designed to assess the HCCV system's performance rather than to make use of an identification. In contrast, when the HCCV system is deployed, the goal is to detect/recognize/classify the individuals it encounters by comparing them to a reference list; being monitored by the HCCV system or being on the reference list thus presents the potential for more acute privacy harms. The collection and use of such images in deployment without informed consent is what directly enables mass surveillance.

Making this distinction between development and deployment has the benefit of enabling HCCV developers to use large corpuses of publicly available images and any other images they collect to train more accurate and less biased HCCV systems. This could promote the creation of larger, fairer publicly available datasets, leveling the playing field for smaller companies.

The drawbacks of this approach are that it does not optimize for individuals maintaining control over their data and, in particular, it does not address the security risks of potential identity theft. These concerns are somewhat mitigated by the fact that copyright protections still apply, so the primary images in question are ones where there is a license available for commercial use. Furthermore, as discussed previously, it is not clear that using images for HCCV creates additional security risks if the images are already publicly available.

Another possible approach would be to make the privacy laws around human images more domain-specific or sectoral. Indeed, federal privacy laws in the US remain sectoral, protecting highly sensitive information in specific contexts, such as medical information.[144] One of the primary sources of imbalance between privacy and fairness considerations in HCCV development is the fact that anti-discrimination protections are highly sectoral, whereas the state biometric privacy protections are not. The innovation of laws like BIPA was to protect specific types of information rather than information in a specific context. While this was motivated by

[142] https://www.scmp.com/comment/opinion/article/3144579/ruling-top-china-court-respects-privacy
[143] https://eare.eu/japan-amends-tdm-exception-copyright/
[144] Daniel J. Solove & Paul M. Schwartz, Information Privacy Law 907 (7th ed.); https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html

the rationale that biometric information is uniquely immutable, this innovation significantly expanded the scope of such laws.

Indeed, even the recent E.U. proposed AI regulation, which has been criticized for being overly broad,[145] focuses specifically on prohibited use, high risk, or limited-risk cases of AI.[146] The regulation provides no requirements for other use cases. Similarly, privacy protections relevant to collecting and processing human images could be limited to contexts like law enforcement, healthcare, finance, employment, education, and any other high-risk domains. This could help distinguish relatively low-risk use cases like unlocking phones, tagging friends on social media, adding filters to faces, etc.

A likely critique of this approach, however, would be that it is difficult to control the contexts in which data are used. Companies often build general-purpose HCCV systems that can be tailored for a wide variety of different domains. Moreover, many concerns about surveillance specifically involve the use of these technologies in relatively low-risk contexts, like tracking people's movements in a mall for ad targeting. The fear is that inferences are being made about people without their knowledge, or that people might self-censor their behavior because of the possibility that they are being watched.[147]

Finally, separate from the privacy protections of the images themselves are the protections around the sensitive attribute data of the image subjects. Making it easier for companies to collect demographic data for the exclusive purpose of conducting audits of their HCCV systems would only narrowly weaken privacy protections while enabling fairer HCCV development. The proposed EU AI regulation gestures in this direction with a carve-out for processing sensitive data for the purposes of complying with other provisions in the regulation.[148] A similar approach could be used in the U.S. to provide carve-outs for algorithmic bias detection and mitigation purposes.

Thus, while any proposals to limit the scope of current privacy protections around HCCV might be highly controversial, making a distinction between data collected for training versus deployment, as well as allowing for sensitive attribute collection and processing for the purposes of addressing algorithmic bias, would help to alleviate some of the challenges practitioners face in developing fair and accurate HCCV systems.

B. Participatory Design

[145] https://www.digitaleurope.org/wp/wp-content/uploads/2021/08/DIGITALEUROPEs-initial-findings-on-the-proposed-AI-Act.pdf; https://www.ft.com/content/a5970b6c-e731-45a7-b75b-721e90e32e1c; https://www2.datainnovation.org/2021-feedback-aia.pdf; https://www.jdsupra.com/legalnews/hogan-lovells-responds-to-the-european-2685952/

[146] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206; https://www.orrick.com/en/Insights/2021/05/The-New-EU-Approach-to-the-Regulation-of-Artificial-Intelligence

[147] http://www.bu.edu/jostl/files/2017/04/Greenbaum-Online.pdf (discussing First Amendment challenges related to FRT). https://repository.law.miami.edu/cgi/viewcontent.cgi?article=1345&context=umlr (broader discussion of importance of privacy in public and anonymity for democratic values).

[148] See supra note []. Title III, Chapter 2, Article 10.5, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

Another approach that scholars in the algorithmic fairness community have proposed is to look toward participatory design—methods that engage stakeholders who use or are affected by a technology in its design[149]—in order to build greater trust between the data subjects and the data collectors. In their piece discussing the parallels between data collection for AI and data collection for archives, Jo and Gebru emphasized the importance of establishing such community relationships and empowering communities to contribute to data collection efforts.[150]

A key difference between archives and datasets for AI, however, is the lack of incentive for most people to contribute to AI datasets. While contributing to an archive can be seen as an honor, a way to preserve the history of your family or community, contributing to an AI dataset is viewed with wariness. Many of the BIPA lawsuits against major US tech companies came after people realized that their Flickr photos were being used in training datasets. An artist even created a platform for people to check whether their images are included in the major publicly available datasets,[151] and journalists wrote of the creepiness of realizing their images were being used.[152]

The challenge for AI companies will be to establish trust with the communities from whom they are collecting images and create incentives for individuals to contribute to dataset collection initiatives. This is easier said than done. For one, the "community" in question might be the global human population if the goal is to ensure that the AI system works well on all people. In addition, community trust will likely be predicated on their images being used only to support HCCV systems that they believe will benefit their communities. The vast majority of data used for training HCCV systems, however, is used to train base models that can perform general tasks—e.g., object, face, or body detection, recognition, and verification—not specific to particular use cases. Companies then adapt these base models to more specific contexts using transfer learning and smaller task- and deployment-specific datasets. Thus, while it might be possible for a company to partner with a specific community to develop an AI system that does a specific trusted task (e.g., a security system for the local school), the base model for such a system would be trained on a large number of photos from other communities. As a result, AI companies typically seek a more global consent for using individuals' photos to develop any computer vision system.

One way, however, to potentially reconcile the desire for both (i) close, carefully designed, and stakeholder-driven data-collection partnerships and (ii) a large breadth of such partnerships is through data consortia and government data collection, which will be discussed in the next Section.

C. Government or Trusted Third-Party Data Collection

[149] https://repositories.lib.utexas.edu/bitstream/handle/2152/28277/SpinuzziTheMethodologyOfParticipatoryDesign.pdf?sequence=2; https://www.researchgate.net/profile/Michael-Muller-12/publication/279063895_Participatory_Design/links/5a82faa1aca272d6501c2deb/Participatory-Design.pdf

[150] https://arxiv.org/pdf/1912.10389.pdf

[151] https://exposing.ai/; https://www.nytimes.com/2021/01/31/technology/facial-recognition-photo-tool.html

[152] https://www.nytimes.com/interactive/2019/10/11/technology/flickr-facial-recognition.html; https://www.nytimes.com/2021/01/31/technology/facial-recognition-photo-tool.html

Another method for addressing these trust issues is to shift the responsibility for data collection and storage from private companies to government agencies and other third-party actors that might be more trusted for data collection. Veale and Ruben, for example, have proposed this approach as a way to handle the privacy concerns around processing sensitive attribute data for bias mitigation.[153] This would have the advantage of creating large image datasets that all companies can use, alleviating some of the challenges facing companies that do not have a pre-existing pipeline for images. In addition, provided that the third-party has strong transparency requirements and governance structures, the data collection process could be more easily evaluated and improved over time, building trust with data subjects. If this entity has sufficient funding and oversight, there should also be a greater incentive for it to uphold high standards and use the latest privacy and bias mitigation techniques.

While this avenue is promising, it is unlikely to resolve all of the major challenges in this space. For HCCV, the problem is not only with the sensitive attribute data used to audit an AI system, but also the fundamental building blocks of the HCCV system itself. Given that HCCV is a highly competitive commercial space, there is a significant incentive for companies to have their own proprietary datasets instead of all using the same government-provided dataset. In addition, while a large image dataset provided by a trusted third-party might be a helpful starting point for training a base model, AI developers still need to collect deployment-context data in order to tailor their models to the particular tasks at hand. For example, an HCCV model that is trying to identify shoplifting needs training data of images or videos of people shoplifting and not shoplifting. These datasets do not need to be as large, but the basic data collection problems discussed above still persist.

Moreover, it is difficult to establish what actor would be sufficiently trustworthy to conduct such large-scale data collection, which could become the basis of many commercial HCCV systems. As mentioned previously, NIST uses mugshots and images of exploited children (among other marginalized populations) as the basis for their test dataset to evaluate commercial facial recognition systems,[154] so we cannot take for granted that government agencies will have easy access to more ethically collected data or that they will enforce the highest standards of informed consent in their data collection practices.

Overall, this approach simply shifts the complications and challenges of ethical data collections discussed above to a new actor. This actor will still have to struggle with questions of how to collect a globally representative dataset with adequate informed consent and sufficiently candid and diverse images. Given that many minority groups might reasonably distrust targeted government data collection efforts, and given the necessity for datasets that feature people and settings from as many countries as possible, government actors will likely struggle to assemble such a dataset. NGOs might similarly struggle to establish trusting relationships with subjects and provide them with sufficient protections around how their data will be used.

Finally, such a solution will likely take many years to develop. There is not necessarily an obvious government agency or NGO who has the technical know-how and national, let alone, international trust to conduct such a wide-scale data collection effort. In the meantime, HCCV technologies will continue to be developed and deployed by companies. Thus, while having

[153] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3060763

[154] https://pages.nist.gov/frvt/reports/1N/frvt_1N_report.pdf; https://www.nist.gov/programs-projects/chexia-face-recognition

trusted third parties collect data is an approach worth exploring, it is unlikely in practice to resolve the core tensions discussed in this Article. At the end of the day, commercial entities will still need to figure out ways to balance these ethical desiderata in their own data collection efforts.

D. Technological Advances

Given the constant advances in HCCV technology, it is important to consider whether the problems addressed in this Article might be resolved over time purely through technological progress. In particular, could advances in privacy-preserving technologies and synthetic image generation address these issues?

Privacy-preserving technologies — including differential privacy, face blurring, and related techniques — are generally helpful for mitigating privacy and security risks with HCCV datasets. Pixelization and blurring are the most well-known techniques, but do not provide any formal privacy guarantee.[155] Differential privacy, in contrast, is a mathematical criterion guaranteeing that the inclusion or exclusion of an individual cannot be distinguished, ensuring that individual-level information will not be leaked.[156] In practice, differential privacy is achieved by adding random noise from a carefully chosen distribution to the data.[157] These techniques, however, do not address the fundamental informed consent problem. For example, if you collect a large dataset of images from the internet, and then you use an algorithm to transform the faces or bodies to be less recognizable, you might still be processing biometric information without informed consent, creating a catch-22. Moreover, there is always a trade-off between the level of privacy attained through such techniques and the utility of the data.[158]

Synthetic image generation is promising in that it can be used to generate images of people who are not real or of real people in new positions/settings, thus augmenting the training dataset. There are two general categories of synthetic image generation approaches relevant to this discussion: ones trained on images of real people and ones that are not. The former, which includes GANs, can modify specific features of an individual (e.g., skin tone, hair length, or perceived gender)[159] or "hallucinate" new people.[160] The models underlying these techniques, however, need to be trained on large numbers of human images, thus undermining the extent to which this approach can resolve the informed consent barrier. These approaches can still be promising, however, as a way to augment existing datasets that have appropriate informed consent.

The second category of approaches often stems from techniques used in animation, like 3D models or CGI. Some of these approaches do not rely on images of real people and instead use drawings, toy models, or 3D computer renderings. These approaches can directly circumvent the need for informed consent as long as images of real people are not processed. The primary

[155] https://arxiv.org/pdf/2102.11072.pdf
[156] https://privacytools.seas.harvard.edu/differential-privacy;
https://privacytools.seas.harvard.edu/files/privacytools/files/pedagogical-document-dp_new.pdf
[157] https://people.csail.mit.edu/asmith/PS/sensitivity-tcc-final.pdf
[158] https://arxiv.org/pdf/2102.11072.pdf
[159] https://arxiv.org/pdf/2007.06570.pdf; https://arxiv.org/pdf/1810.00471.pdf
[160] https://ieeexplore.ieee.org/document/7985521

downsides to this type of approach are i) the difficulties of creating large numbers of highly realistic and diverse images and ii) potential biases of the humans generating these images. The first concern will likely be mitigated over time with advances in this type of technology, propelled by the demand for ever-more realistic-looking fantastical movies. The second issue is more complicated to address. It is inevitable that the people creating these images will have preconceptions of what are relevant types of people and contexts to feature, which has been a critique of start-ups in this space.[161] Creating a sufficiently diverse dataset to reflect the wide array of images an HCCV model is likely to encounter in the real world is a fundamentally challenging problem, even if you have the ability to create realistic images from scratch. Over time, these issues might be mitigated by engaging with diverse image creators and figuring out better ways to measure and audit image datasets for diversity, but for now this is still an open area for future research.

Aside from technologies that side-step or reduce the need for large numbers of human images, federated learning can also be beneficial for giving individuals more control over their data. Federated learning enables data across different parties to be used for training a model, without directly sharing that data between parties.[162] The data remains on the edge device or local server, where a local model is trained and then integrated into the larger model. This is beneficial in contexts where individuals consent to having their images used for training HCCV but are uncomfortable with directly sharing their images with the entity in question. Take, for example, someone is comfortable with their photos being used for training Apple's facial recognition model, but they do not want to directly share their photos with Apple out of concern over how else their photos might be used. Federated learning thus does not solve the fundamental issue that people might not want their images used for training HCCV, but for people who are supportive of the goal of supplying more diverse images to enable training better-performing, less biased HCCV, federated learning can ease some concerns around sharing their data.

Thus, the tension between privacy and fairness in HCCV data collection might be mitigated in the medium- to long-term by technological advances. For now, however, current techniques still rely largely on images of real people and there remain fundamental unsolved questions around how to generate large numbers of diverse, realistic images without substantial bias.

E. Right Against Being Mis-Seen

The final approach is to increase the protections against being "mis-seen." As discussed above in Section V, currently there are only legal protections if being mis-seen triggers a separate legally cognizable harm. As a result, harms that manifest themselves as everyday inconveniences or indignities are unlikely to be protected against, even if the amalgamation of these harms leads to individuals living their lives like second-class citizens. While the general argument of this Article is that there needs to be greater protections against being "mis-seen," this Section will explore possible instantiations of what a right *not* to be "mis-seen" might look like.

[161] https://venturebeat.com/2021/07/06/ai-experts-refute-cvedias-claim-its-synthetic-data-eliminates-bias/

[162] https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

An initial inquiry is whether existing product liability law might be able to provide sufficient protection against being mis-seen by HCCV systems. After all, the harms of being mis-seen are caused by poor product performance, either for everyone or a specific subgroup. Unfortunately, there are several limitations to existing product liability doctrine that would render it unable to provide sufficient protections.

First, product liability law protects primarily against physical harm. If robots and autonomous vehicles become much more widely used in the future, there might be more risk of physical harm from HCCV systems, but for now such systems are primarily deployed in contexts where the potential for bodily harm is minimal (e.g., verifying someone's identity for security purposes, surveilling people, or providing entertainment on social media). While there is the potential to recover for emotional distress under product liability in cases where a bystander is distressed by witnessing a product physically harming another individual,[163] someone experiencing physical harm is still necessary.

A second limitation is that product liability law would not help plaintiffs who experienced algorithmic bias.[164] In cases where the product performs very well for the vast majority of people but poorly on particular subgroups, it would be difficult to establish that the product is unreasonably dangerous.[165] This is especially the case if the HCCV system still performed somewhat well for the subgroups despite a large gap in how it performed across groups.

A third limitation is the lack of robust standards for performance in the AI industry. Current HCCV systems are unlikely to be considered sufficiently inherently dangerous to trigger strict liability, so a negligence standard would likely apply. Industry standards are often relied upon in product liability law to evaluate whether a company has been negligent.[166] While NIST has created a Facial Recognition Vendor Test for companies to benchmark their facial recognition technologies,[167] there is no industry-wide consensus on a single benchmark for performance or what levels of performance are sufficient. Moreover, the need to tailor AI systems to specific deployment contexts suggests that any blanket benchmark or performance standard would be misleading.[168] For example, establishing that a facial recognition system performs well at matching mugshots does not imply it would work well at matching a driver's license photo with a surveillance camera image of a suspect. Surveillance camera images are typically much grainer and lower quality and rarely feature a clear frontal image of the suspect looking into the camera.

While consumer expectations are also often used as a benchmark for reasonableness, as a relatively new but rapidly evolving technology, consumer expectations for HCCV are particularly unstable.[169] This lack of clear consumer expectations has also made it easy for AI

[163] https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=2508

[164] Selbst discusses the challenges of applying negligence law to contexts where there are unevenly distributed harms. https://www.bu.edu/bulawreview/files/2020/09/SELBST.pdf

[165] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3350508

[166] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3350508

[167] https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt

[168] https://lawcat.berkeley.edu/record/1137216/files/01_Solow-Niederman_Web%5B1%5D.pdf. Notably, popular datasets like Labelled Faces in the Wild specifically warn that they should not be used for concluding whether an algorithm would be suitable for commercial purposes, citing lack of diversity along age, gender, ethnicity, lighting conditions, poses, occlusions, and photo resolution. http://vis-www.cs.umass.edu/lfw/.

[169] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3350508

technologies to proliferate while providing minimal representations and warranties to consumers.[170] AI companies often avoid providing any detail about how their technologies are developed or how well they perform on any standardized tests.[171]

Finally, there are many reasonable justifications for why companies do not do more to ensure their computer vision products are not highly biased, making it difficult to pursue a negligence case. As discussed above, both privacy and antidiscrimination laws discourage the collection of data that could be used to test the performance of the AI system across different demographic groups and improve such performance. Current industry practices around preventing algorithmic bias are often minimal due to the lack of incentives to address this issue.[172]

Thus, new laws or regulations are likely needed to protect individuals against being "mis-seen." A right against being "mis-seen" would imply either a private right of action or government audits of HCCV systems. This right could be a general right for HCCV systems to have a minimum performance level, or an anti-discrimination right for the system to not have a significantly disproportionate performance for your subgroup. The former would be most related to negligence and product liability law. As discussed above, establishing standards for reasonableness in the HCCV context might be difficult in the short-term, so strengthening such a right might require the government developing specific HCCV regulations. Transparency obligations could further enable individuals to challenge the use of HCCV systems with poor performance.

The anti-discrimination right would be a new protection that acknowledges the fact that HCCV is increasingly pervasive and embedded into everyday life, creating the risk that those who are more likely to be mis-seen by such technology might find themselves living in a world not optimized for them. Given that HCCV systems lack intentionality,[173] the protection would be against disparate impact, a form of unintentional discrimination whereby facially neutral practices lead to disproportionate adverse effects on particular subgroups. While most anti-discrimination laws apply to specific domains, like employment, finance, or education, this protection would apply to a category of technology, HCCV. While the domain-specificity of many anti-discrimination laws is motivated by the high-stakes nature of those contexts, there are also more anti-discrimination laws like Title II and Title III of the Civil Rights Act of 1964 that protect individuals in low-stakes but commonplace contexts like public accommodation.[174] In addition, the Americans with Disabilities Act of 1960 created accessibility and reasonable accommodation requirements to make it easier for individuals with disabilities to access public services and employment.

[170] https://journals.sagepub.com/doi/abs/10.1177/0022242920953847

[171] Selbst discusses the importance of secrecy in AI development. https://www.bu.edu/bulawreview/files/2020/09/SELBST.pdf

[172] *See* Andrus et al., supra note [].

[173] This is not to say that intentional discrimination on the part of algorithmic developers does not exist, but the examples of algorithmic bias that have been publicly documented stem from unintentional discrimination, so it is important for protections against being "mis-seen" to prevent unintentional discrimination. If a developer does want to create a discriminatory algorithm, however, it is easy to mask their intentions. http://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf; https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=3972&context=dlj

[174] 42 U.S. Code § 2000a.

Having an anti-discrimination right against disparate impact in being "mis-seen" by HCCV technology would thus provide more incentive for companies to directly address issues of algorithmic bias. Of course, this would not directly solve the informed consent challenge posed by privacy laws, but creating such a right would better balance the ethical trade-offs around data collection. Policymakers would need to directly provide guidance more clearly defining the parameters for ethical data collection.

If this protection were enforced by an agency, then there should be resources allocated to investigating allegations of algorithmic bias and conducting audits. This would be especially helpful since algorithmic bias can be very challenging for individuals to detect on their own. Without a concerted effort to gather information about other consumers' experiences and demographics, individuals cannot distinguish between a shoddy product and a biased one.

If the protection were instead enforced through a private right of action, then transparency requirements would be very helpful for enabling consumers to challenge potentially biased products. Of course, the most helpful information would be about the model's performance across different demographic groups.[175] In the absence of that information, however, the requirements should at least include information about the source and properties of the data, the annotation methods, and the testing procedure.[176]

CONCLUSION

Few technologies are as controversial as HCCV, prompting a flurry of privacy laws and moratoriums to be passed in the past several years. Arguments against HCCV generally center on its ability to facilitate mass surveillance and harm women and minority groups through faulty identifications. Not all HCCV enables mass surveillance, however, and the development of more accurate, less biased HCCV requires huge amounts of data, collected from diverse populations with balanced representations. As a result, efforts to improve the fairness and accuracy of HCCV often collide with efforts to enhance privacy protections.

This is not an insurmountable tension—indeed, this Article discusses many potential approaches to address it—but it is a difficult one that will require attention from policymakers and developers to address. Policymakers will need to consider the incentives that developers have under current laws and whether there are ways to both incentivize and enable more efforts to address algorithmic bias in HCCV. Researchers and developers in the HCCV community will need to direct efforts toward studying potential technical solutions to enable HCCV systems to be developed with maximal accuracy and minimal bias while being trained either on smaller, more carefully collected datasets, or on synthetic datasets. Researchers and developers will also need to focus on sociotechnical strategies for ethical data collection, including developing closer

[175] A requirement for such disparate impact assessments was notably missing in the EU's proposed AI regulation, https://www.brookings.edu/blog/techtank/2021/05/04/machines-learn-that-brussels-writes-the-rules-the-eus-new-ai-regulation/.

[176] See Model Cards for a more in-depth discussion about relevant model-related disclosures for transparency purposes. https://arxiv.org/pdf/1810.03993.pdf; Discussing the importance of early documentation for enabling audits. https://dl.acm.org/doi/pdf/10.1145/3375627.3375852

relationships of trust with the communities they seek to collect data from. There is no silver bullet for enabling more ethical HCCV systems that balances all of the concerns this Article surfaces. Breaking down these challenges and potential solutions, however, is an important first step.

More broadly, this Article provides a starting point for more nuanced debates about the appropriate development and use of HCCV. Implicit in the tensions addressed in this Article is the juxtaposition of the suspicion, anxiety, and fear people have toward HCCV and the strong demand for the services such technology can provide. The strategy of addressing the fears around HCCV exclusively through privacy laws and moratoriums is both over- and under-inclusive, increasing the barriers to developing more accurate and less biased HCCV technologies that bear no relation to mass surveillance while also disincentivizing companies from directly addressing issues of algorithmic bias. Instead, a multi-pronged policy strategy is needed, including potential carve-outs for less risky uses of biometric information, support for trusted third-party data collection initiatives, and greater legal protections against being "mis-seen." Ultimately, we must balance the desire not to be seen with the desire not to be invisible.