Harvard Journal of Law & Technology
Volume 36, Number 1 Fall 2022

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

*Alice Xiang**

ABSTRACT

The rise of facial recognition and related computer vision technologies has been met with growing anxiety over the potential for artificial intelligence ("AI") to create mass surveillance systems and further 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. Reducing bias in human-centric computer vision systems ("HCCV"), including facial recognition, can involve collecting large, diverse, and candid image datasets, which can run counter to privacy protections.

It is intuitive to think that being "unseen" by AI is preferable — that being underrepresented in the data used to develop facial recognition might somehow allow one to evade mass surveillance. As we have seen in the law enforcement context, however, just because facial recognition technologies are less reliable at identifying people of color has not meant that they have not been used to surveil these communities and deprive individuals of their liberty. Thus, being "unseen" by AI does not protect against being "mis-seen." While in the law enforcement context this tension can simply be resolved by prohibiting the use of facial recognition technology, HCCV encompasses a much broader set of technologies, from face detection for a camera's autofocus feature to pedestrian detection on a self-driving car.

The first contribution of this Article is to characterize this tension between privacy and fairness in the context of algorithmic bias mitigation for HCCV. In particular, this Article argues that the irreducible paradox underlying current efforts to design less biased HCCV is the simultaneous desire to be "unseen" yet not "mis-seen" by AI. Second, the Article reviews the strategies 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

* Alice Xiang is the Global Head of AI Ethics for Sony Group and the Lead Research Scientist for AI ethics at Sony AI. Thanks to Dan Ho, participants of the 2021 We Robot conference, and my colleagues at Sony AI for their comments on earlier drafts. I also thank Rui-Jie Yew for her research assistance and the editors of the *Harvard Journal of Law & Technology* for their edits and support.

data, using privacy-preserving techniques, generating synthetic data, obtaining informed consent, and expanding regulatory mandates or government audits. Finally, this Article argues that solving this paradox requires considering the importance of not being "mis-seen" by AI rather than simply being "unseen." De-tethering these notions (being seen versus unseen versus mis-seen) can help clarify what rights relevant laws and policies should seek to protect. For example, this Article will examine the implications of a right not to be disproportionately mis-seen by AI, in contrast to regulations around what data should remain unseen. Given that privacy and fairness are both critical objectives for ethical AI, it is vital for lawmakers and technologists to address this tension head-on; approaches that rely purely on visibility or invisibility will likely fail to achieve either objective.

TABLE OF CONTENTS

I. INTRODUCTION

Human-centric computer vision ("HCCV") technologies,[1] including facial recognition, are some of the most controversial artificial intelligence ("AI") technologies. HCCV systems are among the few types

[1] 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 on which Model Cards focuses. Margaret Mitchell et al., *Model Cards for Model Reporting*, PROC. ACM CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 220, 220 (2019) https://dl.acm.org/doi/10.1145/3287560.3287596 [https://perma.cc/2696-5XHJ]. 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." Inioluwa Deborah Raji & Genevieve Fried, *About Face: A Survey of Facial Recognition Evaluation*, AAAI 2020 WORKSHOP ON AI EVALUATION 1, 2 (2021). HCCV includes tasks involving human bodies and objects. HCCV can also be seen as any computer vision system relying on "people-centric" datasets. Margot Hanley, Apoorv Khandelwal, Hadar Averbuch-Elor, Noah Snavely & Helen Nissenbaum, *An Ethical Highlighter for People-Centric Dataset Creation*, NAVIGATING THE BROADER IMPACTS OF AI RSCH. WORKSHOP AT NEURIPS 1, 1–2 (2020), https://arxiv.org/pdf/2011.13583.pdf [https://perma.cc/6ZQR-ASHS].

of AI that have been subject to bans or moratoriums. Many U.S. jurisdictions have restricted the use of facial recognition technologies ("FRT") by government entities, particularly law enforcement.[2] The current version of the proposed EU AI Act categorizes all remote biometric identification ("RBI") systems as high-risk (and thus subject to extensive regulatory requirements),[3] and prohibits the use of real-time RBI by law enforcement in public spaces (with some narrow carveouts).[4] From a privacy perspective, the specter of mass surveillance, particularly by state actors, has led to significant criticism of the growing pervasiveness of FRT[5] and growing pushes for strengthening information privacy laws.

In recent years, there has also been a growing awareness of the issues of bias in HCCV. The highly influential *Gender Shades* paper showed that many of the major commercial gender classification algorithms performed less effectively on women than men and less well on individuals with deeper skin tones than lighter skin tones.[6] Since then, subsequent studies, including one by the National Institute of Standards and Technology ("NIST"), part of the U.S. Department of Commerce, have shown differences in performance based on skin tone and gender

[2] *See, e.g.*, Elec. Priv. Info. Ctr., *State Facial Recognition Policy*, EPIC.ORG, https://epic.org/state-policy/facialrecognition [https://perma.cc/LC5L-GUFU] (listing moratoriums or bans in California and Massachusetts); Grace Woodruff, *Maine Now Has the Toughest Facial Recognition Restrictions in the U.S.*, SLATE (July 2, 2021, 5:50 AM), https://slate.com/technology/2021/07/maine-facial-recognition-government-use-law.html [https://perma.cc/NK5F-J829] (describing Maine's ban); *Facial Recognition Technology Ban Passed by King County Council*, KING CNTY. (June 1, 2021), https://kingcounty.gov/council/mainnews/2021/June/6-01-facial-recognition.aspx [https://perma.cc/4HWV-4CRJ] (describing King County's ban in Washington state).

[3] *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, at 4, COM (2021) 206 final (Apr. 21, 2021) [hereinafter *Proposed EU AI Act*] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206 [https://perma.cc/X5E2-ZWMX].

[4] *Id.*

[5] *See, e.g.*, Antoaneta Roussi, *Resisting the Rise of Facial Recognition*, 587 NATURE 350, 351 (2020); *Facial Recognition & Biometric Mass Surveillance: Document Pool*, EUR. DIGIT. RTS. (Mar. 25, 2020), https://edri.org/our-work/facial-recognition-document-pool [https://perma.cc/X6L4-BAQS]; *Ban Dangerous Facial Recognition Technology That Amplifies Racist Policing*, AMNESTY INT'L (Jan. 26, 2021), https://www.amnesty.org/en/latest/news/2021/01/ban-dangerous-facial-recognition-technology-that-amplifies-racist-policing [https://perma.cc/U4Y3-HEK3]; *Face Recognition Technology*, ACLU, https://www.aclu.org/issues/privacy-technology/surveillance-technologies/face-recognition-technology [https://perma.cc/TF42-VRZJ].

[6] Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. 1ST CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY: PROC. MACH. LEARNING RSCH. 1, 8 (2018).

for different HCCV systems.[7] These studies have attributed these biases to a lack of diversity in the datasets used to train these commercial AI systems.[8]

Simultaneously addressing these concerns around privacy and fairness is difficult in practice. To address bias in HCCV, researchers at IBM created the "Diversity in Faces" ("DiF") dataset,[9] which was initially received positively for being far more diverse and balanced than previous face image datasets.[10] DiF, however, soon became embroiled in controversy once journalists highlighted the fact that the dataset consisted of images from Flickr.[11] The Flickr images had Creative Commons licenses, covering the copyright of the images, but the plaintiffs had not consented to having their images used in facial recognition training datasets.[12] In part due to this controversy, IBM announced it would be discontinuing its facial recognition program.[13] Microsoft, Amazon, and Google, which also used the DiF dataset to improve the fairness of their models, were also sued.[14] This example highlights a core tension in developing less biased HCCV: We want AI to recognize us, but we are uncomfortable with the idea of AI having access to data about us. While creating large, diverse human image datasets with informed consent is not impossible (as Section V.A discusses), there are challenges that require further research and regulatory guidance.

This tension is further amplified when the need for sensitive attribute data is considered. For example, 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 often discomfiting and raise further privacy

[7] PATRICK GROTHER, MEI NGAN & KAYEE HANAOKA, U.S. DEP'T OF COM., NAT'L INST. OF STANDARDS & TECH. INTERAGENCY OR INTERNAL REP. 8280, FACE RECOGNITION VENDOR TEST (FRVT) PART 3: DEMOGRAPHIC EFFECTS 2 (2019), https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf [https://perma.cc/U9EW-FYDH].

[8] *Id.*; *see* Buolamwini & Gebru, *supra* note 6, at 88.

[9] Michele Merler, Nalini Ratha, Rogerio Feris & John R. Smith, *Diversity in Faces*, IBM RSCH. AI 1, 1 (2019), https://arxiv.org/pdf/1901.10436.pdf [https://perma.cc/VA69-8SQL].

[10] Kyle Wiggers, *IBM Releases Diversity in Faces, a Dataset of Over 1 Million Annotations to Help Reduce Facial Recognition Bias*, VENTUREBEAT (Jan. 29, 2019, 5:59 AM), https://venturebeat.com/2019/01/29/ibm-releases-diversity-in-faces-a-dataset-of-over-1-million-annotations-to-help-reduce-facial-recognition-bias [https://perma.cc/6Q4U-6C85].

[11] Stephen Shankland, *IBM Stirs Controversy by Using Flickr Photos for AI Facial Recognition*, CNET (Mar. 13, 2019, 1:07 PM), https://www.cnet.com/news/ibm-stirs-controversy-by-sharing-photos-for-ai-facial-recognition [https://perma.cc/J452-KZNK].

[12] Taylor Hatmaker, *Lawsuits Allege Microsoft, Amazon and Google Violated Illinois Facial Recognition Privacy Law*, TECHCRUNCH (July 15, 2020, 4:59 PM), https://techcrunch.com/2020/07/15/facial-recognition-lawsuit-vance-janecyk-bipa [https://perma.cc/LA5F-H7HX].

[13] Nicolas Rivero, *The Influential Project that Sparked the End of IBM's Facial Recognition Program*, QUARTZ (July 20, 2022), https://qz.com/1866848/why-ibm-abandoned-its-facial-recognition-program [https://perma.cc/M3YK-S8HB].

[14] *See* Hatmaker, *supra* note 12.

concerns. In designing DiF, the researchers did not have a variable for race or ethnicity, so they used various computational methods to derive labels for different facial features to indirectly capture differences across race, including metrics for skin color and craniofacial areas.[15] While these features were used in an effort to ensure racial diversity without access to direct data on race, these approaches do not reflect 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.[16]

Other attempts at creating diverse face image datasets, like FairFace,[17] approach the challenge by having "Mechanical Turkers" ("MTurkers")[18] guess people's demographic attributes. If at least two MTurkers agree, then the label is considered ground truth; if there is no agreement, the image is discarded.[19] This approach is concerning because 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 process 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 laws address this issue primarily by erring on the side of hiding people's personal data unless there is explicit informed consent. In fact, privacy law and antidiscrimination law are often viewed as symbiotic,[20] under the assumption that preventing companies from collecting personal information helps to prevent discrimination. Evidence of bias in FRT, however, has contradicted this notion. Low representation of minority groups in the datasets used to train such models has led to biased performance, but that has not prevented the use of such systems to deprive minority individuals of their liberty. There have been several cases of Black men in the United States who were wrongfully arrested due to faulty facial recognition matches.[21] In

[15] *See* Merler et al., *supra* note 9, at 3.

[16] *See* Blaise Agüera y Arcas, Margaret Mitchell & Alexander Todorov, *Physiognomy's New Clothes*, MEDIUM (May 6, 2017), https://medium.com/@blaisea/physiognomys-new-clothes-f2d4b59fdd6a [https://perma.cc/KLC3-E4R2].

[17] Kimmo Kärkkäinen & Jungseock Joo, *FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation*, PROC. IEEE WINTER CONF. ON APPLICATIONS COMPUT. VISION 1547, 1550 (2021).

[18] *Id.* Amazon Mechanical Turk is a crowdsourcing platform where developers can put tasks for crowd workers (also known as MTurkers) to complete for payment. AMAZON MECH. TURK, https://www.mturk.com [https://perma.cc/3GVX-4YE4].

[19] Kärkkäinen & Joo, *supra* note 17.

[20] Jessica L. Roberts, *Protecting Privacy to Prevent Discrimination*, 56 WM. & MARY L. REV. 2097, 2112 (2015).

[21] *See, e.g.*, Kashmir Hill, *Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match*, N.Y. TIMES (Jan. 6, 2021), https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html [https://perma.cc/G9M6-FJLC]; Kashmir Hill, *Wrongfully*

2019, for example, Nijeer Parks, a Black man, was arrested due to a faulty facial recognition match; he spent ten days in jail and paid close to $5,000 to defend himself before the case was dismissed for lack of evidence.[22]

To address such issues of bias in FRT, the policy response has centered around moratoriums on the usage of FRT by law enforcement and other public agencies.[23] While such moratoriums are reasonable given current problems with such technologies, they are limited to specific jurisdictions, do not apply to other domains for FRT, and do not address bias in other forms of HCCV.[24] The lack of stronger regulatory incentives to address bias in HCCV is concerning given the growing use of such technology. While there is limited data about the broader HCCV market, the FRT market alone is projected to grow from $4.45 billion in 2021 to $12.11 billion in 2028.[25] Even in North America, where FRT has been quite controversial, the market for FRT is expected to double by 2027.[26] While recent moratoriums on FRT for law enforcement suggest a strong discomfort with government use of the technology, the demand for private surveillance camera systems with FRT has continued to grow,[27] as has the use of this technology in everyday life. It is

Accused by an Algorithm, N.Y. TIMES (Aug. 3, 2020), https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html [https://perma.cc/VU8K-BY82]; Elisha Anderson, *Controversial Detroit Facial Recognition Got Him Arrested for a Crime He Didn't Commit*, DET. FREE PRESS (July 11, 2020, 11:03 PM), https://www.freep.com/story/news/local/michigan/detroit/2020/07/10/facial-recognition-detroit-michael-oliver-robert-williams/5392166002 [https://perma.cc/TXV4-SD63].

[22] *See* Hill, *Another Arrest*, *supra* note 21.

[23] *See supra* note 2 and accompanying text.

[24] For example, California enacted a moratorium on FRT use in police body cameras, while Somerville, MA, and Oakland, CA, enacted bans of FRT use by city agencies, including police departments. *See* Elec. Priv. Info. Ctr., *supra* note 2. These bans are limited to these specific jurisdictions, only apply to specific government uses of FRT, and do not go beyond FRT to address other forms of HCCV. *Id.*

[25] GRAND VIEW RSCH., FACIAL RECOGNITION MARKET SIZE, SHARE & TRENDS ANALYSIS REPORT BY TECHNOLOGY (2D, 3D, FACIAL ANALYTICS), BY APPLICATION (ACCESS CONTROL, SECURITY & SURVEILLANCE), BY END-USE, BY REGION, AND SEGMENT FORECASTS, 2021-2028 (2021), https://www.grandviewresearch.com/industry-analysis/facial-recognition-market [https://perma.cc/G9BZ-LMCA].

[26] *In Charts: Facial Recognition Technology — and How Much Do We Trust It?*, FIN. TIMES (May 16, 2021), https://www.ft.com/content/f6a9548a-a235-414e-b5e5-3e262e386722 [https://perma.cc/9RRA-6UW6].

[27] *See, e.g.*, Lance Whitney, *Demand for Video Surveillance Cameras Expected to Skyrocket*, TECHREPUBLIC (July 14, 2020, 8:32 AM), https://www.techrepublic.com/article/demand-for-video-surveillance-cameras-expected-to-skyrocket [https://perma.cc/TJ4R-SW6L]; Lauren Bridges, *Amazon's Ring Is the Largest Civilian Surveillance Network the U.S. Has Ever Seen*, GUARDIAN (May 18, 2021, 8:51 AM), https://www.theguardian.com/commentisfree/2021/may/18/amazon-ring-largest-civilian-surveillance-network-us [https://perma.cc/HV4H-JD3U]; Edvardas Mikalauskas, *The Rise of the Private Surveillance Industry*, CYBERNEWS (Feb. 15, 2022), https://cybernews.com/privacy/the-rise-of-the-private-surveillance-industry [https://perma.cc/7V88-EBXS]; *Global Video Surveillance Market Revenues to Exceed $24B by End of 2021*, CEPRO (Aug. 9, 2021), https://www.cepro.com/security/global-video-surveillance-market-revenues-exceed-24b-2021 [https://perma.cc/X3MZ-EZSM].

now common for people to open their phones with face verification[28] or to automatically sort photos based on the people in the photos.[29] Moreover, in regions where FRT has not faced as much controversy as in the United States or EU,[30] the technology is increasingly used by government authorities[31] and for everyday verification purposes, such as payment[32] and entering establishments.[33] Outside of FRT, the use of

[28] *See* Luana Pascu, *Biometric Facial Recognition Hardware Present in 90% of Smartphones by 2024*, BiometricUpdate.com (Jan. 7, 2020, 1:52 PM), https://www.biometricupdate.com/202001/biometric-facial-recognition-hardware-present-in-90-of-smartphones-by-2024 [https://perma.cc/KW73-GRJC].

[29] *See Find and Identify Photos of People Using Photos on Mac*, Apple, https://support.apple.com/guide/photos/view-photos-by-whos-in-them-phtad9d981ab/mac [https://perma.cc/X6Q2-5CAV]; *Search by People, Things & Places in Your Photos*, Google Photos Help, https://support.google.com/photos/answer/6128838?hl=en&co=GENIE.Platform%3DAndroid [https://perma.cc/CRE3-2YXJ].

[30] *See* Léa Steinacker, Miriam Meckel, Genia Kostka & Damian Borth, *Facial Recognition: A Cross-National Survey on Public Acceptance, Privacy, and Discrimination*, 37 Proc. Int'l Conf. on Mach. Learning, Law & Mach. Learning Workshop 1, 4 (2020), https://arxiv.org/pdf/2008.07275.pdf [https://perma.cc/8FTC-SEPL]. 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. *See* Eva Dou, *China Built the World's Largest Facial Recognition System. Now, It's Getting Camera Shy*, Wash. Post (July 30, 2021, 2:56 AM), https://www.washingtonpost.com/world/facial-recognition-china-tech-data/2021/07/30/404c2e96-f049-11eb-81b2-9b7061a582d8_story.html [https://perma.cc/43HN-XCH2]; Sam Shead, *Chinese Residents Worry About Rise of Facial Recognition*, BBC (Dec. 5, 2019), https://www.bbc.com/news/technology-50674909 [https://perma.cc/QL69-728Q]; Stella Yifan Xie, *In China, Paying With Your Face Is Hard Sell*, Wall St. J. (Sept. 20, 2020, 6:20 AM), https://www.wsj.com/articles/in-china-paying-with-your-face-is-hard-sell-11600597240 [https://perma.cc/B39Y-PG2X]. That said, the use of facial recognition is far more pervasive in China than in the United States. Paul Mozur, *Inside China's Dystopian Dreams: A.I., Shame and Lots of Cameras*, N.Y. Times (July 8, 2018), https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html [https://perma.cc/2JXQ-PK7X].

[31] *See, e.g.*, Aloysius Low, *In Singapore, Facial Recognition Is Getting Woven into Everyday Life*, NBC News (Oct. 12, 2020, 12:04 PM), https://www.nbcnews.com/tech/tech-news/singapore-facial-recognition-getting-woven-everyday-life-n1242945 [https://perma.cc/PPY6-TZYU]; Dave Davies, *Facial Recognition and Beyond: Journalist Ventures Inside China's 'Surveillance State,'* NPR (Jan. 5, 2021, 12:50 PM), https://www.npr.org/2021/01/05/953515627/facial-recognition-and-beyond-journalist-ventures-inside-chinas-surveillance-sta [https://perma.cc/VH4G-52KV].

[32] *See* Richard Baimbridge, *Why Your Face Could Be Set to Replace Your Bank Card*, BBC (Jan. 24, 2021), https://www.bbc.com/news/business-55748964 [https://perma.cc/M4BG-4W4X]; *4 Japan Firms to Tie Up in Facial Recognition for Payment*, Jiji Press Eng. News Serv. (Aug. 2, 2021), https://www.japantimes.co.jp/news/2021/08/02/business/tech/facial-recognition-tie-up [https://perma.cc/SZ4M-PXQG]; Leeloo Tang, *Revisiting Facial-Recognition Payment: Old Problems Still Lingering*, Nielsen Norman Grp. (Mar. 20, 2022), https://www.nngroup.com/articles/facial-recognition-payment [https://perma.cc/D9UL-3R32] ("[M]ore than 495 million Chinese used facial-recognition payment in 2021 — that's roughly one third of China's population.").

[33] *See* Tang, *supra* note 32 ("For example, nowadays many residential buildings are equipped with face-recognition devices at entrances."); Kosuke Shimizu, Ryosuke Hanada & Takashi Kawakami, *Japan in Race with China for Facial-Recognition Supremacy*, Nikkei Asia (Dec. 20, 2019), https://asia.nikkei.com/Business/Business-trends/Japan-in-race-with-China-for-facial-recognition-supremacy [https://perma.cc/G7KY-RH48]; Chris Gallagher, *Masks No Obstacle for New NEC Facial Recognition System*, Reuters (Jan. 7, 2021, 12:39 AM), https://www.reuters.com/article/us-health-coronavirus-japan-facial-recog/masks-no-

HCCV has become increasingly common, with cameras using eye, face, or body detection for autofocus[34] or for creating artificial bokeh effects (blurry background),[35] robots using human/object detection to navigate real-world spaces,[36] and computer-generated images ("CGI") employing AI to create new fantastical scenes.[37]

Thus, while privacy and bias concerns around FRT have manifested themselves in moratoriums on specific use cases in some jurisdictions, HCCV systems are unlikely to go away anytime soon. The focus of this Article is therefore not on the line-drawing exercise of which HCCV systems should be banned or permitted but rather on categorizing the relevant harms from such technologies and tensions that must be addressed first to judge what would constitute appropriate use cases. Policymakers and AI developers must assess 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 both 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, where addressing bias can be enabled by additional access to personal information. The goal of this Article is not to advocate for the increased or decreased usage of HCCV technologies but rather to characterize this tension between privacy and bias mitigation efforts and to propose potential paths forward that respect both.

Part II lays out important definitions used throughout the Article to facilitate more nuanced discourse about HCCV. Part III discusses the importance of considering the harms of being "mis-seen" in a world where HCCV is increasingly pervasive. Part IV explains what makes the HCCV context unique in terms of the privacy and fairness tensions

obstacle-for-new-nec-facial-recognition-system-idUSKBN29C0JZ [https://perma.cc/3DUG-YZVD].

[34] *See, e.g.*, Gaudenz Boesch, *Face Detection: Real-time Applications with Deep Learning (2022 Guide)*, VISO.AI, https://viso.ai/deep-learning/face-detection-overview [https://perma.cc/6XCE-UA9W]; Matthew Richards & Marcus Hawkins, *Exploring Canon's Intelligent Autofocus System*, CANON, https://www.canon-europe.com/pro/stories/intelligent-autofocus-explained [https://perma.cc/GN6K-X424].

[35] *See, e.g.*, Matic Broz, *Luminar AI: Portrait Bokeh AI*, PHOTUTORIAL (Apr. 17, 2022), https://photutorial.com/luminar-ai-bokeh-ai [https://perma.cc/2F6S-XUSE].

[36] *See, e.g.*, William P. Shackleford, Geraldine Cheok, Tsai H. Hong, Kamel Saidi & Michael O. Shneier, *Performance Evaluation of Human Detection Systems for Robot Safety*, 83 J. INTELLIGENT & ROBOTIC SYS. 85, 90 (2016), https://www.nist.gov/publications/performance-evaluation-human-detection-systems-robot-safety [https://perma.cc/SW77-AFP8]; Nicola Bellotto & Huosheng Hu, *Multisensor-Based Human Detection & Tracking for Mobile Service Robots*, 39 IEEE TRANSACTIONS ON SYS., MAN & CYBERNETICS 167, 167 (2009).

[37] *See, e.g.*, Dushyant Mehta et al., *XNect: Real-Time Multi-Person 3D Motion Capture with a Single RGB Camera*, 39 PROC. ACM TRANSACTIONS ON GRAPHICS 1, 1–2 (2020); *Human Pose Estimation with Deep Learning: Overview for 2022*, SUPERANNOTATE (Feb. 24, 2022), https://blog.superannotate.com/human-pose-estimation-with-deep-learning [https://perma.cc/XHX9-5X6J].

it raises. Part V discusses current challenges to mitigating algorithmic bias in HCCV, focusing particularly on the difficulties with collecting large, diverse datasets with informed consent. Part VI discusses relevant privacy laws in the United States and EU. Part VII elaborates on the harms associated with being "seen" versus "mis-seen." Part VIII evaluates potential solutions for better balancing protections against being "seen" versus "mis-seen."

II. DEFINITIONS

Resolving the tensions between two very important ethical desiderata — privacy and fairness — requires a nuanced approach. The discourse around HCCV rarely distinguishes between the many types of technologies implicated, which differ widely in terms of their potential societal harms. This Part will seek to provide the vocabulary needed for greater nuance by clarifying technologies and concepts that are often conflated.

Throughout this Article, I will use the term HCCV to refer specifically to AI systems that rely on images of humans for their training and test data.[38] These are the AI technologies whose *development* is directly affected by biometric information privacy regulations that protect information extracted from human faces or bodies. I stress the word "development" because the human images I address in this Article are the images in the training set used to teach the HCCV system how to detect, recognize, classify, or extract features from people or objects or the images in the evaluation sets used to evaluate the model's performance. These images used for development are typically distinct from the images the HCCV system perceives in deployment.[39] The question of which images should be used when the system is deployed is inextricably tied to the highly context-specific exercise of determining which use cases of HCCV should be permitted versus banned. Although this issue is a highly important policy question, it is beyond the scope of this Article. For example, whether police should be allowed to run drivers' license photos through an FRT to identify potential sus-

[38] *See supra* note 1 and accompanying text for discussion of related terms in the existing literature.

[39] The images used for development are existing images collected by developers in datasets. In many deployment contexts, the images being perceived are new images. For example, a human detection model deployed in a shopping mall to count the number of customers at any given time will be perceiving images of the mall's activity in real time. There are edge cases where an image in deployment was used in training. For example, if you use a facial recognition model to identify a person in a preexisting image online, it is possible that the image might have been used by the model's developer for training. Another possible exception is if images perceived in deployment are collected and later used to retrain the model. This Article does not encourage expanding the deployment of HCCV solely for gathering more diverse data for future training.

pects cannot be separated from the question of whether this is an appropriate use case, whereas many of the large image datasets used to develop the FRT are not specific to particular use cases.[40]

In characterizing the privacy concerns around HCCV, this Article will focus on biometric information privacy risks given that this is the primary area where there has been regulation and litigation around the images used to develop HCCV systems.[41] That said, even if images featuring biometric information are not involved, there might still be legal issues with copyright and privacy concerns if photos taken inside people's homes are used. Note that certain medical use cases of computer vision (e.g., melanoma detection) might or might not count as HCCV for the purposes of this Article's discussion depending on whether the images used to develop the AI include faces or hands of the individuals.[42]

The primary computer vision tasks motivating this piece are face detection, verification, identification, and analysis, 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 can also be relevant if developers use images featuring both people and objects to train their models.

Although many HCCV technologies are often colloquially referred to as "facial recognition technologies," FRT is only a small subset of HCCV. For the purposes of this Article, HCCV encompasses all computer vision technologies whose development requires human-centric personal information — thus confronting current information privacy laws. In addressing the tensions between existing privacy laws and HCCV bias mitigation efforts, it is thus important to note that HCCV

[40] Given the expenses required to collect large image datasets from scratch and the large amounts of computation to train such computer vision models, researchers commonly train use-case specific models using transfer learning and fine-tuning. *See Fine-Tuning*, DIVE INTO DEEP LEARNING, https://d2l.ai/chapter_computer-vision/fine-tuning.html [https://perma.cc/2F44-QQ87]. This means that a source model trained on a large number of images for a more general task (e.g., ImageNet has 10 million images and is commonly used to train object recognition source models) is used as a starting point and then trained with additional data for the target task (e.g., chair recognition). *Id.* The intuition is that even if the source dataset had little to do with the target task, the source model has learned many general image features, like edges, textures, shapes, and object composition, that are useful for more specific tasks. *See id.*; *Transfer Learning and Fine-Tuning*, TENSORFLOW, https://www.tensorflow.org/guide/keras/transfer_learning [https://perma.cc/4PTD-3H9E].

[41] *See, e.g.*, Rui-Jie Yew & Alice Xiang, *Regulating Facial Processing Technologies: Tensions Between Legal and Technical Considerations in the Application of Illinois BIPA*, PROC. ACM CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 1017, 1017 (2022).

[42] Illinois's Biometric Information Privacy Act ("BIPA"), for example, defines "biometric identifier[s]" as "retina or iris scan, fingerprint, voiceprint, or scan of hand or face geometry." 740 ILL. COMP. STAT. 14/10 (2008) [hereinafter BIPA]. Typically, images used for training melanoma detection models are close-ups of the skin, so biometric information privacy laws would not apply. *See, e.g.*, Veronica Rotemberg et al., *A Patient-Centric Dataset of Images and Metadata for Identifying Melanomas Using Clinical Context*, 8 SCI. DATA, Jan. 28, 2021, at 1.

includes technologies, as enumerated below, that largely do not figure in policy conversations about privacy laws.[43] Note that the paragraphs below do not seek to classify these technologies into "acceptable" versus "unacceptable" bins, but rather to illustrate the wide variety of HCCV technologies.

Face detection involves detecting whether a human face is in an image and, if so, returning the location and extent of the face (typically through drawing a bounding box).[44] This is one of the most frequently used face-related computer vision tasks and serves as the basis for the other face-related tasks (one must first detect a face before one can identify or analyze it).[45] Face or body detection is often used to count people or to trigger a subsequent task. For example, an AI-assisted elevator might count the number of people in the elevator and not stop for additional people if the elevator is at capacity.[46]

Face verification and identification are related tasks for identifying a person. Face verification refers to a one-to-one comparison between a reference face and a new face.[47] When unlocking a phone, a face verification algorithm is used to compare the face perceived by the camera with the reference face for the owner of the phone.[48] Facial identification refers to one-to-many comparisons; the perceived face is compared against a reference set of faces to identify which (if any) of the reference faces is a match.[49] If police have an image of a suspect, a face identification system could compare the image to a reference set of

[43] For example, EFF, a major civil liberties organization advocating for user privacy, specifically separates face detection from FRT as "not rais[ing] significant privacy concerns." Adam Schwartz, Nathan Sheard & Bennett Cyphers, *Face Recognition Technology: Commonly Used Terms*, ELEC. FRONTIER FOUND. (Oct. 7, 2021), https://www.eff.org/deeplinks/2021/10/face-recognition-technology-commonly-used-terms [https://perma.cc/D6U8-LGYD].

[44] *See* Ming-Hsuan Yang, David J. Kriegman & Narendra Ahuja, *Detecting Faces in Images: A Survey*, 24 IEEE TRANSACTIONS ON PATTERN ANALYSIS & MACH. INTEL. 1, 1 (2002); Jason Brownlee, *How to Perform Face Detection with Deep Learning*, MACH. LEARNING MASTERY (June 3, 2019), https://machinelearningmastery.com/how-to-perform-face-detection-with-classical-and-deep-learning-methods-in-python-with-keras [https://perma.cc/LW7V-GX9R].

[45] *See* Shervin Minaee, Ping Luo, Zhe Lin & Kevin Bowyer, *Going Deeper into Face Detection: A Survey*, ARXIV, Apr. 13, 2021, at 1, https://arxiv.org/abs/2103.14983 [https://perma.cc/N4PZ-ZNYF]; Samuel Dooley, George Z. Wei, Tom Goldstein & John P. Dickerson, *Robustness Disparities in Face Detection*, PROC. CONF. ON NEURAL INFO. PROCESSING SYS. 1, 1 (2022).

[46] N. V. Rajeesh Kumar, G. DhanaSekar & M. Dennis, *Application of Face Detection System for Passenger Counting in Lifts Using Haar Features*, 11 ARPN J. ENG'G & APPLIED SCIS. 8336, 8336–37 (2016).

[47] *See* DOUGLAS YEUNG, REBECCA BALEBAKO, CARLOS IGNACIO GUTIERREZ & MICHAEL CHAYKOWSKY, RAND CORP., FACE RECOGNITION TECHNOLOGIES: DESIGNING SYSTEMS THAT PROTECT PRIVACY AND PREVENT BIAS 8–9 (2020).

[48] *See* JOY BUOLAMWINI, VICENTE ORDÓÑEZ, JAMIE MORGENSTERN & ERIK LEARNED-MILLER, FACIAL RECOGNITION TECHNOLOGIES: A PRIMER 5 (2020).

[49] *See id.* at 10.

driver's license photos to see if there is a match. Face identification can also be used in social media applications to generate tag suggestions.[50]

Face analysis, or "face attribute classification,"[51] refers to the task of automatically generating labels for a face.[52] For example, the model might label faces as "male" or "female." This type of task can be fraught from an ethical perspective, given concerns about how much information can be accurately discerned from someone's face. Gender classification has especially been criticized since gender cannot be assessed purely based on a photo, especially if an individual is transgender or non-binary.[53] In addition, controversial technologies like emotion recognition and 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., someone might smile through discomfort or sadness.[54] In addition, efforts to use face analysis to identify who might be a better job candidate or who might have a propensity for criminal behavior have been highly criticized as pseudoscientific.[55] That said, face 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.[56] AI-assisted medical analyses of a person's body or face can also fall into this category.

There are also common HCCV tasks that focus on the entire body rather than just the face. Body detection, for example, might be used by an autonomous vehicle to detect and avoid pedestrians.[57] Pose estimation is used to estimate the spatial key points of a person's joints to

[50] Face identification is often colloquially referred to as "facial recognition." *See, e.g.*, Queenie Wong, *Facebook Replaces Setting That Only Suggested Friends to Tag in Photos*, CNET (Sept. 3, 2019, 1:19 PM), https://www.cnet.com/news/privacy/facebook-replaces-setting-that-only-suggested-friends-to-tag-in-photos [https://perma.cc/369A-52SQ].

[51] *See* BUOLAMWINI ET AL., *supra* note 48, at 6.

[52] *See, e.g.*, Schwartz et al., *supra* note 43.

[53] *See* Os Keyes, *The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition*, 2 PROC. ACM CONF. ON HUM.-COMPUT. INTERACTION 1, 1 (2018).

[54] *See* Douglas Heaven, *Expression of Doubt*, 578 NATURE 502, 503 (2020); Madhumita Murgia, *Emotion Recognition: Can AI Detect Human Feelings from a Face?*, FIN. TIMES (May 12, 2021), https://www.ft.com/content/c0b03d1d-f72f-48a8-b342-b4a926109452 [https://perma.cc/6BZT-XMVE]; Kate Crawford, *Artificial Intelligence Is Misreading Human Emotion*, THE ATLANTIC (Apr. 27, 2021), https://www.theatlantic.com/technology/archive/2021/04/artificial-intelligence-misreading-human-emotion/618696 [https://perma.cc/M86S-PQPM].

[55] Agüera y Arcas et al., *supra* note 16; *Facial Recognition to 'Predict Criminals' Sparks Row Over AI Bias*, BBC (June 24, 2020), https://www.bbc.com/news/technology-53165286 [https://perma.cc/YE4C-GDUR]; Jeremy Kahn, *HireVue Drops Facial Monitoring Amid A.I. Algorithm Audit*, FORTUNE (Jan. 19, 2021, 12:01 PM), https://fortune.com/2021/01/19/hirevue-drops-facial-monitoring-amid-a-i-algorithm-audit [https://perma.cc/X5LK-34QX].

[56] Katherine Boehret, *New Cameras Guarantee a Smile on Your Face*, WALL ST. J. (Apr. 23, 2008, 12:01 AM), https://www.wsj.com/articles/SB120889435178135615 [https://perma.cc/YW3W-P52G].

[57] *See, e.g.*, Kirti Balani, Sneha Deshpande, Ranjit Nair & Vishal Rane, *Human Detection for Autonomous Vehicles*, PROC. IEEE INT'L TRANSP. ELECTRIFICATION CONF. 1, 1 (2015);

determine whether an individual is doing a certain activity.[58] In a security context, the goal might be to detect whether someone is shoplifting.[59] Such technologies are also commonly used for augmented reality or CGI.[60] Pose estimation typically does not involve identifying the person, but it can be used for such purposes. For example, 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 United States and EU.[61]

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, these technologies are often trained on images featuring humans. For example, in the research community, the Common Objects in Context ("COCO") dataset is one of the most commonly used datasets for object-related tasks.[62] This dataset features around two hundred thousand images with humans and objects labeled. Using images with humans can be helpful given that, in the real world, we are often interested in detecting objects that humans are interacting with. Training an object recognition model exclusively on images without humans might make it more difficult for the model to perform well

Benjamin Wilson, Judy Hoffman & Jamie Morgenstern, *Predictive Inequity in Object Detection*, ARXIV, Feb. 21, 2019, at 1, https://arxiv.org/pdf/1902.11097.pdf [https://perma.cc/64DA-BZMV].

[58] *See, e.g.*, Leonid Sigal, *Human Pose Estimation*, *in* COMPUT. VISION 362, 362, 368 (Katsushi Ikeuchi ed., 2016), https://www.cs.ubc.ca/~lsigal/Publications/SigalEncyclopediaCVdraft.pdf [https://perma.cc/X5AU-7MVG].

[59] *See, e.g.*, Yohei Katabuchi, *Artificial Intelligence Becomes "Shoplifting G-men" Damage Amount Reduced By 40%*, ITMEDIA (May 28, 2018, 4:00 PM), https://www.itmedia.co.jp/news/articles/1805/28/news085.html [https://perma.cc/B3KX-DYN7]; Pranav Dar, *'AI Guardman' — A Machine Learning Application That Uses Pose Estimation to Detect Shoplifters*, ANALYTICS VIDHYA (June 27, 2018), https://www.analyticsvidhya.com/blog/2018/06/ai-guardman-machine-learning-application-estimates-poses-detect-shoplifters [https://perma.cc/2Y7Y-PJG2]; Guillermo A. Martínez-Mascorro, José R. Abreu-Pederzini, José C. Ortiz-Bayliss & Hugo Terashima-Marín, *Suspicious Behavior Detection on Shoplifting Cases for Crime Prevention by Using 3D Convolutional Neural Networks*, 9 COMPUTATION 1, 3 (2020), https://arxiv.org/pdf/2005.02142.pdf [https://perma.cc/9DEP-3AWH].

[60] *See* Mehta et al., *supra* note 37.

[61] *Summary of the Opinion of the European Data Protection Supervisor on the European Commission's White Paper on Artificial Intelligence — A European Approach to Excellence and Trust*, 2020 O.J. (C 392) 3, 3; California Consumer Privacy Act of 2018, Cal. Civ. Code § 1798.140; Consolidated Laws of New York, Chapter 57-A, Art. 1, § 106-B; Dan Cooper & Gemma Nash, *UK ICO Publishes New Guidance on Special Category Data*, COVINGTON (Nov. 29, 2019), https://www.insideprivacy.com/eu-data-protection/uk-ico-publishes-new-guidance-on-special-category-data [https://perma.cc/Q6NN-W8UW].

[62] COMMON OBJECTS IN CONTEXT DATASET, https://cocodataset.org [https://perma.cc/LT4M-UDN6].

in real-world contexts.[63] Often, the goal is not object recognition in isolation but to combine human detection with object recognition. An airport, for example, might want to detect abandoned luggage for security purposes.[64]

Thus, although FRT has 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 is subject to biometric information privacy laws, raising difficult questions around privacy and fairness.

For more general terms, 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."[65] "Model" is a more precise term, but "algorithm" is more commonly used in colloquial discussions about AI. In general, I will use colloquial terms when referencing popular discourse that uses those terms. 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 encompasses computer vision contexts where there are privacy considerations under existing biometric information privacy laws, which will be discussed in Part VI. Being "unseen" thus means not having your images or images of people like you collected or processed for developing HCCV (i.e., included in training, validation, or test sets). This includes images used to train the base model, which performs a more general task, and images collected in the specific domain for the specific task. Note that being "seen"/"unseen" focuses specifically on 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, since those images are generally not useful for improv-

[63] Note that recent research has shown that using privacy-preserving techniques like face blurring can enable object recognition to be trained on such datasets while reducing the privacy risk. Kaiyu Yang, Jacqueline Yau, Li Fei-Fei, Jia Deng & Olga Russakovsky, *A Study of Face Obfuscation in ImageNet*, 39 PROC. INT'L CONF. ON MACH. LEARNING 1, 1 (2022), https://arxiv.org/abs/2103.06191 [https://perma.cc/K67T-RFAV]. Face blurring will be discussed further in Section VIII.D.

[64] *See generally* Jing-Ying Chang, Huei-Hung Liao & Liang-Gee Chen, *Localized Detection of Abandoned Luggage*, EURASIP J. ON ADVANCES SIGNAL PROCESSING, June 2010, at 1.

[65] *Algorithm*, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/algorithm#note-1 [https://perma.cc/HZC2-FQ95].

ing the fairness of the model, so there is usually no fairness versus privacy tension in deployment. Note that this distinction can break down in contexts where the model continues to learn from images collected in deployment. In such contexts, the status quo prioritization of privacy is reasonable.[66]

Being "mis-seen" refers to experiencing poor performance from a deployed HCCV system: this includes your face/body not being detected, being misrecognized as someone else, someone else being misrecognized for you, or having images/videos of you misclassified or mischaracterized. This last category includes tasks like suspicious behavior detection, where you might be erroneously labeled as cheating on an exam[67] or shoplifting.[68] As will be explored in greater depth in Part VII, 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.

The tension between not wanting to be "seen" or "mis-seen" resembles the tension between "visibility" and "hypervisibility" in that it centers on the challenges of increasing representation in a way that is not harmful to the marginalized individuals represented.[69] These dichotomies are distinct, however, in that having one's images included in a dataset to train or test an HCCV model does not necessarily have the implications of hypervisibility in terms of heightened scrutiny or surveillance. Scrutiny or surveillance by HCCV comes at the point of deployment rather than development. Being included versus excluded in the training and test sets used to develop the model affects the accuracy of the model on individuals like you. It does not necessarily affect whether the system will be deployed on individuals like you or whether you will be included in a reference set. The excessive deployment of such technologies to surveil marginalized communities is what leads to these problems of hypervisibility. Moreover, whereas hypervisibility is often associated with tokenization and distorted visibility — the tendency to provide visibility disproportionately to negative representations of marginalized individuals[70] — algorithmic bias mitigation

[66] This approach is uncommon in deployment, however, given that it requires someone to label the new images collected to continue training the model.

[67] *See* Kashmir Hill, *Accused of Cheating by an Algorithm, and a Professor She Had Never Met*, N.Y. TIMES (May 27, 2022), https://www.nytimes.com/2022/05/27/technology/college-students-cheating-software-honorlock.html [https://perma.cc/L83M-UWH2].

[68] *See supra* note 59.

[69] *See generally* Isis H. Settles, NiCole T. Buchanan & Kristie Dotson, *Scrutinized but Not Recognized: (In)visibility and Hypervisibility Experiences of Faculty of Color*, 113 J. VOCATIONAL BEHAV. 62 (2019).

[70] *See generally* Rasul A. Mowatt, Bryana H. French & Dominique A. Malebranche, *Black/Female/Body Hypervisibility and Invisibility: A Black Feminist Augmentation of Feminist Leisure Research*, 45 J. LEISURE RSCH. 644 (2013).

efforts in HCCV often revolve precisely around preventing models from learning stereotyped representations of people (e.g., that only men play outdoor sports), as will be discussed further in Part IV and Section VII.B. Thus, while the visibility versus hypervisibility tension is highly relevant to the discourse around whether, how, where, and for whom HCCV systems should be deployed, it is only indirectly related to the tension between being "seen" versus "mis-seen."

Lastly, it is important to define the term "bias." Because "bias" is a catch-all term for many different types of disparities, it is beneficial to consider more explicitly the specific harms involved.[71] In this Article, I will use the term "bias" to refer to the disparate performance of the HCCV system (e.g., different rates of misrecognition, misdetection, or misclassification) across different groups that might lead to disproportionate harm for specific groups. In Part VII, I break down the specific types of bias harms associated with being "mis-seen." "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 while preserving privacy.

III. Why Worry About Being Mis-Seen?

Given that this Article focuses on the current tensions and imbalances between privacy and fairness when developing HCCV, it is important to address the basic question of why being "mis-seen" is such a problem. While being "seen" by an HCCV system without informed consent is considered, under some privacy laws, to be a harm in and of itself, being "mis-seen" is only considered to be harmful if it leads to a separate legally cognizable harm. For example, when Robert Williams sued the Detroit Police Department after a faulty facial recognition match, he brought his action under the Fourth Amendment right to be free from unlawful seizures and the Elliot-Larsen Civil Rights Act.[72]

One could argue that this distinction is reasonable — that the harms of being mis-seen are already appropriately accounted for through existing antidiscrimination laws and other laws. Antidiscrimination law,

[71] *See* Solon Barocas et al., *Designing Disaggregated Evaluations of AI Systems: Choices, Considerations, and Tradeoffs*, AAAI/ACM CONF. ON AI, ETHICS, & SOC'Y 368, 375 (2021); *see also* Su Lin Blodgett, Solon Barocas, Hal Daumé III & Hanna Wallach, *Language (Technology) Is Power: A Critical Survey of "Bias" in NLP*, 58 PROC. ANN. MEETING ASS'N FOR COMPUTATIONAL LINGUISTICS 5454, 5454 (2020) (delineating specific harms).

[72] *Farmington Hills Father Sues Detroit Police Department for Wrongful Arrest Based on Faulty Facial Recognition Technology*, ACLU MICH. (Apr. 13, 2021), https://www.aclumich.org/en/press-releases/farmington-hills-father-sues-detroit-police-department-wrongful-arrest-based-faulty [https://perma.cc/Q9RP-NK8F]; *see also* MICH. COMP. LAWS § 37.2202 (1977) (protecting against government entities denying an individual "full and equal utilization of . . . public service[s] . . . because of . . . race").

however, primarily applies in specific, comparatively high-stakes contexts like employment,[73] housing,[74] finance,[75] and the public sector.[76] While the limited domains of antidiscrimination law might be reasonable in the context of discrimination by human actors, algorithmic discrimination raises additional concerns. The growing proliferation of HCCV in everyday life suggests that even small or subtle biases might accumulate into substantial harm.

Imagine, for example, being an individual of an underrepresented 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,[77] but it does not recognize you, so you must 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.[78] 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.[79] 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 workday, 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.[80] Sitting in the darkness, you are confronted with your own invisibility.

[73] *See, e.g.*, Title VII of the Civil Rights Act of 1964, 42 U.S.C. §§ 2000e–2000e-17.

[74] *See, e.g.*, Fair Housing Act, Title VIII of the Civil Rights Act of 1968, 42 U.S.C. §§ 3601–3619.

[75] *See, e.g.*, Equal Credit Opportunity Act, 15 U.S.C. § 1691.

[76] *See, e.g.*, Title VI of the Civil Rights Act of 1964, 42 U.S.C. § 2000d–2000d-7.

[77] Facial recognition payment is increasingly commonly used, particularly in East Asia. *See* Tang, *supra* note 32.

[78] *Id.*

[79] Facial recognition for entering buildings is also very common and allows individuals to enter or exit without a physical key or fob. *See supra* note 33.

[80] AI is increasingly being used for HVAC systems. *See* Alex Makarevich, *Why AI Technology for HVAC Is the Next Big Thing in the Commercial Real Estate Market*, SOFTEQ (Jan. 4, 2022), https://www.softeq.com/blog/why-ai-technology-for-hvac-is-the-next-big-thing-in-the-commercial-real-estate-market [https://perma.cc/D2V2-6PJN] ("The AI technology is capable of quick adjustments to occupancy conditions. Depending on the number of people in the building, the system turns on/off air conditioning, regulates the temperature accordingly,

In this hypothetical scenario, featuring currently extant technologies, I have only discussed a few of the possible 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.

Of course, the scenario 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. In the absence of antidiscrimination protections, however, the primary forces preventing poor performance are the competitiveness of the market and the desire of companies to produce high-performing products. Such incentives might be insufficient if the system works very well for most people; those in the minority group might be seen as out-of-distribution edge cases that do not need to be specifically addressed. There is no legal protection for the individual in this scenario.

IV. 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[81] and in analyses of relevant antidiscrimination law prohibitions on the usage of such data.[82] This Article 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* the data used in developing such models. In tabular or language data contexts, stripping the dataset of personally identifiable information ("PII") can significantly mitigate privacy risks.[83] In contrast, for

modifies other settings, and keeps bills low."); *see also* Gregory Barber, *Energy-Saving AI Is Coming for Your Office Thermostat*, Wired (Apr. 1, 2020, 6:00 AM), https://www.wired.com/story/energy-saving-ai-controls-lights-office-thermostat [https://perma.cc/ERE6-4BRH].

[81] *See, e.g.*, McKane Andrus, Elena Spitzer, Jeffrey Brown & Alice Xiang, *"What We Can't Measure, We Can't Understand": Challenges to Demographic Data Procurement in the Pursuit of Fairness*, Proc. ACM Conf. on Fairness, Accountability, & Transparency 249, 250 (2021).

[82] Alice Xiang, *Reconciling Legal and Technical Approaches to Algorithmic Bias*, 88 Tenn. L. Rev. 649, 666 (2021).

[83] For tabular data, removing unique identifiers, employing differential privacy techniques, and limiting the number 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.

HCCV, even if the developer strips all the metadata, the face or body images themselves can constitute PII under certain laws.[84] Moreover, developing HCCV usually involves the processing of biometric information, which is subject to additional privacy protections, as will be discussed further in Part VI. Section VIII.D will discuss in greater depth the potential utility of face blurring and other image anonymization techniques, but, in short, depending on the type of HCCV being developed, such techniques cannot guarantee that the person cannot be identified while still preserving the ability to train an accurate model.

Not only are the privacy concerns stronger in the HCCV context, but the need for wide-ranging data collection efforts is also greater. While a simple logistic regression model with tabular data can be trained on thousands of instances, HCCV can require millions of images to train a model to perform certain recognition tasks.[85] Moreover, while dataset diversity is important in all contexts, bias in computer vision is particularly strongly connected with a lack of sufficient dataset diversity. Much of the literature on algorithmic bias in computer vision attributes such biases to a lack of sufficiently large, diverse, and balanced datasets.[86] This is not to say that data bias is the only source of algorithmic bias[87] or that improving the data used in development is the only way to reduce bias in models,[88] but it is a key factor.

[84] ERIKA MCCALLISTER, TIM GRANCE & KAREN SCARFONE, U.S. DEP'T OF COM., NAT'L INST. OF STANDARDS & TECH., GUIDE TO PROTECTING THE CONFIDENTIALITY OF PERSONALLY IDENTIFIABLE INFORMATION (PII) 2-1 (2010).

[85] Note, however, that it is fine for some subset of these images to be synthesized. Iacopo Masi, Anh Tuấn Trần, Tal Hassner, Jatuporn Toy Leksut & Gérard Medioni, *Do We Really Need to Collect Millions of Faces for Effective Face Recognition?*, 14 PROC. EUR. CONF. ON COMPUT. VISION 579, 580 (2016); *see also* Kate Conger, *How Apple Says It Prevented Face ID from Being Racist*, GIZMODO (Oct. 16, 2017, 5:56 PM), https://gizmodo.com/how-apple-says-it-prevented-face-id-from-being-racist-1819557448 [https://perma.cc/Z2ML-MFQW] (describing Apple using over a billion images to train Face ID).

[86] *See* Angelina Wang et al., *REVISE: A Tool for Measuring and Mitigating Bias in Visual Datasets*, 130 INT'L J. COMPUT. VISION 1790 (2022); Buolamwini & Gebru, *supra* note 6, at 1–2; GROTHER ET AL., *supra* note 7, at 1–2 (2019).

[87] Modeling choices such as the architecture, loss function, optimizer, and hyperparameters can also affect the fairness of a model. *See* Sara Hooker, *Moving Beyond "Algorithmic Bias Is a Data Problem,"* 2 PATTERNS 1, 3 (2021).

[88] For example, for image captioning models, approaches have been proposed that would focus the model's attention on relevant features instead of irrelevant ones correlated with demographics. *See, e.g.*, Lisa Anne Hendricks, Kaylee Burns, Kate Saenko, Trevor Darrell & Anna Rohrbach, *Women Also Snowboard: Overcoming Bias in Captioning Models*, 15 PROC. EUR. CONF. ON COMPUT. VISION 793, 794–95 (2018); Zeyu Wang et al., *Towards Fairness in Visual Recognition: Effective Strategies for Bias Mitigation*, IEEE/CVF CONF. ON COMPUT. VISION & PATTERN RECOGNITION 8916, 8916 (2020). There are also bias mitigation approaches using synthetic data. *See, e.g.*, Guha Balakrishnan, Yuanjun Xiong, Wei Xia & Pietro Perona, *Towards Causal Benchmarking of Bias in Face Analysis Algorithms*, 16 PROC. EUR. CONF. ON COMPUT. VISION 547, 549 (2020) https://dl.acm.org/doi/abs/10.1007/978-3-030-58523-5_32 [https://perma.cc/5XPY-9MZL]; Erroll Wood, Tadas Baltrušaitis, Charlie Hewitt, Sebastian Dziadzio, Thomas J. Cashman & Jamie Shotton, *Fake It Till You Make It: Facial Analysis in the Wild Using Synthetic Data Alone*, PROC. IEEE/CVF INT'L CONF. ON COMPUT. VISION 3681, 3688 (2021).

In contrast, in the tabular data context, collecting data on more diverse individuals rarely solves issues of algorithmic bias. For example, in criminal justice data in the United States, there is evidence that Black individuals have faced higher rates of arrest for drug-related crimes despite similar rates of drug offenses.[89] Algorithmic risk assessment tools 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 contribution of these biases toward higher arrest rates and counteract those biases in the data (e.g., through algorithmic rebalancing across groups[90] or trying to find less biased features for predicting criminal offense rather than arrest[91]).

Algorithmic bias in the HCCV context generally boils down to two problems: (1) lack of representation[92] and (2) spurious correlations.[93] 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 facial recognition, where studies have shown that people have a harder time recognizing people of other races.[94] 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 when they are growing up.[95] Similar to humans, facial recognition models also exhibit an "other-race effect," whereby algorithms developed in

[89] *See, e.g.*, Sharad Goel, Justin M. Rao & Ravi Shroff, *Precinct or Prejudice? Understanding Racial Disparities in New York City's Stop-and-Frisk Policy*, 10 ANNALS OF APPLIED STAT. 365, 365 (2016); Kristian Lum & William Isaac, *To Predict and Serve?*, SIGNIFICANCE, Oct. 2016, at 14, 18; Emma Pierson et al., *A Large-Scale Analysis of Racial Disparities in Police Stops Across the United States*, 4 NATURE HUM. BEHAV. 736, 736 (2020).

[90] *See generally* SOLON BAROCAS, MORITZ HARDT & ARVIND NARAYANAN, FAIRNESS AND MACHINE LEARNING: LIMITATIONS AND OPPORTUNITIES (2022) (discussing possible approaches for achieving balance across different fairness metrics). Such methods can be preprocessing, in-processing, or post-processing methods.

[91] Riccardo Fogliato, Alice Xiang, Zachary Lipton, Daniel Nagin & Alexandra Chouldechova, *On the Validity of Arrest as a Proxy for Offense: Race and the Likelihood of Arrest for Violent Crimes*, PROC. AAAI/ACM CONF. ON AI, ETHICS, & SOC'Y 100, 104 (2021).

[92] *See, e.g.*, Buolamwini & Gebru, *supra* note 6, at 2; Terrance DeVries, Ishan Misra, Changshan Wang & Laurens van der Maaten, *Does Object Recognition Work for Everyone?*, PROC. IEEE/CVF CONF. ON COMPUT. VISION & PATTERN RECOGNITION WORKSHOPS 52, 52 (2019).

[93] *See, e.g.*, Robert Geirhos et al., *Shortcut Learning in Deep Neural Networks*, 2 NATURE MACH. INTEL. 665, 670 (2020); Hendricks et al., *supra* note 88, at 793.

[94] *See* Agata Blaszczak-Boxe, *Some People Suffer from Face Blindness for Other Races*, SCI. AM. MIND (May 1, 2017), https://www.scientificamerican.com/article/some-people-suffer-from-face-blindness-for-other-races [https://perma.cc/R2PX-XZX9].

[95] 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. Elinor McKone et al., *A Critical Period for Faces: Other-Race Face Recognition Is Improved by Childhood but Not Adult Social Contact*, SCI. REPS., Sept. 6, 2019, at 1, 1.

Western countries perform better for Caucasian faces and algorithms developed in East Asian countries perform better for East Asian faces.[96] 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, then 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 contains misleading patterns, often due to societal biases.[97] 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 the reverse for outdoor images,[98] perpetuating long-standing stereotypes of women inhabiting domestic spheres and men inhabiting public spheres. Even though the background of an image should be irrelevant 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 women indoors and men outdoors. Thus, it is important to develop training datasets that are well-balanced to avoid spurious correlations. For example, the proportion of women indoors should roughly match the proportion of men indoors. Of course, it is impossible to account for all possible spurious correlations, so researchers typically focus on ones that are related to pernicious societal stereotypes around attributes like gender, age, race, ethnicity, or skin tone.[99] Collecting a balanced dataset in an unbalanced world, however, is difficult in practice, as the next Part 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.[100] Outside of the HCCV context, bias mitigation itself can be a source of controversy.[101] For example, scholars have explored the ways

[96] P. Jonathon Phillips, Fang Jian, Abhijit Narvekar, Julianne Ayyad & Alice J. O'Toole, *An Other-Race Effect for Face Recognition Algorithms*, 8 ACM TRANSACTIONS ON APPLIED PERCEPTION, Jan. 2011, at 1, 1.

[97] This is related to the problem of shortcut learning. *See* Geirhos et al., *supra* note 93, at 665.

[98] *See* Wang et al., *supra* note 86, at 1798.

[99] *See, e.g.*, Buolamwini & Gebru, *supra* note 6; Balakrishnan et al., *supra* note 88.

[100] 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. Pinar Barlas, Kyriakos Kyriakou, Olivia Guest, Styliani Kleanthous & Jahna Otterbacher, *To "See" Is to Stereotype: Image Tagging Algorithms, Gender Recognition, and the Accuracy-Fairness Trade-off*, PROC. ACM ON HUM.-COMPUT. INTERACTION, Dec. 2020, at 1, 4.

[101] Nicol Turner Lee, Paul Resnick & Genie Barton, *Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms*, BROOKINGS (May 22, 2019), https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-

in which many of the dominant approaches to bias mitigation in the tabular data machine learning context are tantamount to affirmative action and may actually violate antidiscrimination law because of their reliance on quotas, different thresholds, or other forms of rebalancing across the protected attribute.[102] In contrast, ensuring that a model recognizes people based on their facial features and not based on irrelevant characteristics such as the image background is important not only for reducing bias but also for increasing accuracy and robustness across a wider set of deployment contexts.

Thus, there are many aspects of computer vision that make the tensions between privacy and fairness particularly salient and difficult to untangle. That said, many of the insights from this Article are not unique to computer vision. If we can reconcile the tensions between privacy and fairness in HCCV, we might be able to apply analogous solutions to other forms of AI.

V. Challenges to Algorithmic Bias Mitigation in Computer Vision

Collecting larger, more diverse training and test datasets serves two aims: (1) improving the overall accuracy and robustness of the model and (2) mitigating potential biases. While this Article addresses both 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. While accuracy of major commercial facial recognition technologies has improved dramatically over the past few years, issues of bias persist.[103]

Though 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 to collect the large corpora of data needed to train their systems. In the United States, the National Institute of Standards and Technology ("NIST") uses police mugshots

practices-and-policies-to-reduce-consumer-harms [https://perma.cc/Y3J8-5G7E] (discussing fairness-accuracy trade-off).

[102] Such approaches may be considered legally suspect affirmative action or reverse discrimination. Daniel E. Ho & Alice Xiang, *Affirmative Algorithms: The Legal Grounds for Fairness as Awareness*, U. Chi. L. Rev. Online (Oct. 30, 2020), https://lawreviewblog.uchicago.edu/2020/10/30/aa-ho-xiang [https://perma.cc/RQ8C-ZK6E]; Xiang, *supra* note 82, at 654; Jason R. Bent, *Is Algorithmic Affirmative Action Legal?*, 108 Geo. L.J. 803, 808 (2020).

[103] U.S. Gov't Accountability Off., GAO-20-522, Facial Recognition Technology: Privacy and Accuracy Issues Related to Commercial Uses 25 (2020).

and images of exploited children,[104] individuals crossing the border, and visa applicants in its test dataset, which is used by major companies to benchmark the performance of their commercial FRT.[105] In China, 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."[106]

While those datasets were collected by government entities, there are also many large publicly available human image datasets collected by academic or industry researchers. These typically rely on web-scraped photos. Some datasets focus on celebrities or public figures (e.g., MS-Celeb-1M);[107] others focus on a broader array of subjects through online platforms like Flickr (e.g., YFCC100M),[108] which made large numbers of images public and easily downloadable with Creative Commons licenses permitting their use for commercial purposes.

Images of celebrities have especially assisted with the advancement of research into facial recognition and verification systems since such datasets include many images of the same person at different times, angles, and settings. Such datasets, however, raise issues around consent and biases introduced by only training algorithms to recognize celebrities, whose features are not representative of the general population.[109]

[104] These images are used specifically to test the performance of face detection and recognition systems on children. *Chexia Face Recognition*, U.S. DEP'T OF COM., NAT'L INST. OF STANDARDS & TECH. (May 8, 2019), https://www.nist.gov/programs-projects/chexia-face-recognition [https://perma.cc/7BU5-DHYN]. Images of children are hard to come by in most datasets due to additional privacy restrictions.

[105] Os Keyes, Nikki Stevens & Jacqueline Wernimont, *The Government Is Using the Most Vulnerable People to Test Facial Recognition Software*, SLATE (Mar. 17, 2019, 8:32 PM), https://slate.com/technology/2019/03/facial-recognition-nist-verification-testing-data-sets-children-immigrants-consent.html [https://perma.cc/2L6P-3MJH]; PETER GROTHER, MEI NGAN, KAYEE HANAOKA, JOYCE C. YANG & AUSTIN HOM, U.S. DEP'T OF COM., NAT'L INST. OF STANDARDS & TECH., ONGOING FACE RECOGNITION VENDOR TEST (FRVT) PART 1: VERIFICATION 55–56 (2022); PETER GROTHER, MEI NGAN & KAYEE HANAOKA, U.S. DEP'T OF COM., NAT'L INST. OF STANDARDS & TECH., DRAFT SUPPLEMENT OF INTERAGENCY REP. 8271, FACE RECOGNITION VENDOR TEST (FRVT) PART 2: IDENTIFICATION 6 (2022) ("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.").

[106] Paul Mozur, *One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority*, N.Y. TIMES (Apr. 14, 2019), https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html [https://perma.cc/EN7L-MN8F].

[107] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He & Jianfeng Gao, *MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition*, 14 PROC. EUR. CONF. ON COMPUT. VISION 87, 89 (2016) (featuring 10 million face images of nearly 100,000 individuals).

[108] Bart Thomee et al., *YFCC100M: The New Data in Multimedia Research*, 59 COMMC'NS ACM, Feb. 2016, at 64, 66 (featuring around 100 million images and videos).

[109] 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

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 constructed large public datasets using Flickr images were often motivated to use Flickr to address the issues of bias that plague other datasets.[110] Flickr-based datasets feature photos of non-celebrities[111] from amateur photographers,[112] yielding a large amount of diversity.[113] Recently, however, there have been many lawsuits leveraging Illinois's Biometric Information Privacy Act ("BIPA")[114] against companies using such datasets since the individuals in the Flickr images did not consent to having their photos used to train facial recognition algorithms.[115] 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.

smokey eye and lipstick). *See* Jungseock Joo & Kimmo Kärkkäinen, *Gender Slopes: Counterfactual Fairness for Computer Vision Models by Attribute Manipulation*, 2 PROC. INT'L WORKSHOP ON FAIRNESS, ACCOUNTABILITY, TRANSPARENCY & ETHICS MULTIMEDIA 3 (2020). 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, and white woman. Vinay Prabhu, Dian Ang Yap, Alexander Wang & John Whaley, *Covering Up Bias in CelebA-Like Datasets with Markov Blankets: A Post-Hoc Cure for Attribute Prior Avoidance*, WORKSHOP ON INVERTIBLE NEURAL NETS & NORMALIZING FLOWS, 2019, at 1, 1.

[110] Aaron Nech & Ira Kemelmacher-Shlizerman, *Level Playing Field for Million Scale Face Recognition*, PROC. IEEE CONF. ON COMPUT. VISION & PATTERN RECOGNITION 3406, 3406 (2017); *see also* Merler et al., *supra* note 9, at 7; Tsung-Yi Lin et al., *Microsoft COCO: Common Objects in Context*, 13 PROC. EUR. CONF. ON COMPUT. VISION 740, 745 (2014).

[111] *See* Nech & Kemelmacher-Shlizerman, *supra* note 110.

[112] *See* Lin et al., *supra* note 110.

[113] *See* Merler et al., *supra* note 9, at 7.

[114] 740 ILL. COMP. STAT. 14/15 (2008).

[115] Olivia Solon, *Facial Recognition's 'Dirty Little Secret': Millions of Online Photos Scraped Without Consent*, NBC (Mar. 17, 2019, 11:25 AM) https://www.nbcnews.com/tech/internet/facial-recognition-s-dirty-little-secret-millions-online-photos-scraped-n981921 [https://perma.cc/5VLW-NXG3]; Sara Merken, *IBM Can't Shake Facial Recognition Suit but Dodges Some Claims*, REUTERS LEGAL (Sept. 16, 2020, 8:10 PM), https://www.reuters.com/article/dataprivacy-ibm-biometrics/ibm-cant-shake-facial-recognition-suit-but-dodges-some-claims-idUSL1N2GD2JP [https://perma.cc/7VLH-NVNE]; Taylor Hatmaker, *supra* note 12.

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 in their face-tagging algorithm.[116] That said, for companies with a business model where individuals upload large numbers of diverse photos, there is the possibility of directly asking users for consent.

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 their photos being used for facial recognition, the consent of the other individuals in the user's photos is still necessary. Even if the individuals have social media accounts where they have provided approval on their end, it is unclear how the social media platform can know whether the individuals in the photo have given consent without first attempting to recognize the individuals. Moreover, depending on the company's privacy policy, the images collected through the platform may or may not be eligible for use in developing HCCV.[117]

For academic researchers, public sector entities, or companies without business models that incentivize organic data collection, the need to collect large, diverse datasets with informed consent poses additional difficulties. Companies can buy images from vendors that work with crowd workers who upload images of themselves to the platform in return for payment,[118] but it is difficult to (1) obtain enough data and (2) obtain sufficiently diverse and candid data. While social media companies do 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. Each image can cost several dollars,[119]

[116] Order re Final Approval, Attorneys' Fees and Costs, and Incentive Awards at 1, *In re* Facebook Biometric Info. Priv. Litig., 522 F. Supp. 3d 617 (N.D. Cal. Feb. 26, 2021) (No. 15-cv-03747-JD).

[117] *See, e.g.*, Sam Shead, *Facebook Trains A.I. to 'See' Using 1 Billion Public Instagram Photos*, CNBC (Mar. 4, 2021, 1:52 PM), https://www.cnbc.com/2021/03/04/facebook-trains-ai-to-see-using-1-billion-public-instagram-photos-.html [https://perma.cc/JLG9-HQ2V] (describing how while some Instagram users "may be surprised to hear that their images are being used to train Facebook AI systems," Instagram's data policy includes using such information for research and innovation).

[118] *See, e.g.*, *Crowd Capabilities*, APPEN, https://appen.com/crowd-2/#capabilities [https://perma.cc/GV3L-8KN4]; *Data Collection & Creation Services*, TELUS INT'L, https://www.telusinternational.com/solutions/ai-data-solutions/data-collection-services [https://perma.cc/2PQ3-HACB]; *Reliable AI Data Collection Services to Train ML Models*, SHAIP, https://www.shaip.com/offerings/data-collection [https://perma.cc/J6GP-G2BR].

[119] *See* Gerard Andrews, *What Is Synthetic Data?*, NVIDIA BLOG (June 8, 2021), https://blogs.nvidia.com/blog/2021/06/08/what-is-synthetic-data [https://perma.cc/TVB2-TDHV] ("A single image . . . could cost $6 from a labeling service."). In the author's own experience obtaining quotes from real image vendors, the costs can be as high as $40 per image, depending on the diversity and annotation requirements.

so training an HCCV model with strong performance from scratch using images procured this way can cost millions or even billions of dollars. In explaining how they addressed potential issues of bias, Apple cited using over a billion images of diverse individuals to train Face ID, their face verification system.[120] In addition, Facebook used over 1 billion public images from Instagram in developing its object recognition model SEER.[121]

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 in this Article.[122] There are very few companies that have the advantage of a large, global, diverse user base willing to upload billions of images for free. There are far more companies, academics, and public sector entities that either operate or seek to operate in the HCCV space.

In addition, when crowd 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, one photo holding an object, one photo sitting/standing/running, one photo occluded by an object), the photos generally look staged.[123] In computer vision, "in the wild" is a phrase used to refer 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 within.[124] When buying photos from crowd workers, however, it can be difficult to collect large numbers of unconstrained images.

[120] Conger, *supra* note 85.

[121] Shead, *supra* note 117; Goyal et al., *Self-Supervised Pretraining of Visual Features in the Wild*, ARXIV, Mar. 5, 2021, at 1, https://arxiv.org/abs/2103.01988 [https://perma.cc/5PLG-ZG2D].

[122] Competition concerns have also motivated the creation of large publicly available web-scraped datasets. *See* Kashmir Hill & Aaron Krolik, *How Photos of Your Kids Are Powering Surveillance Technology*, N.Y. TIMES (Oct. 11, 2019), https://www.nytimes.com/interactive/2019/10/11/technology/flickr-facial-recognition.html [https://perma.cc/LQ92-TWT7] ("The database creators said their motivation was to even the playing field in machine learning. Researchers need enormous amounts of data to train their algorithms, and workers at just a few information-rich companies — like Facebook and Google — had a big advantage over everyone else.").

[123] 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. Raji & Fried, *supra* note 1, at 2. 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.

[124] Gary B. Huang, Marwan Mattar, Tamara Berg & Eric Learned-Miller, *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*, DANS WORKSHOP ON FACES 'REAL-LIFE' IMAGES: DETECTION, ALIGNMENT & RECOGNITION, Oct. 2008, at 1, 2–3.

These challenges create several performance and bias concerns. First, an HCCV model trained on heavily 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, expressions, or occlusions (e.g., hats, masks, or sunglasses).[125] 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 — that can exacerbate 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. For example, research has shown that object recognition models trained predominantly on U.S. data struggle to accurately recognize common objects like soap and cooking equipment in developing country contexts.[126]

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 (e.g., perceived gender, age, and ethnicity), hairstyles, clothing, lighting conditions, background, pose, and camera type.[127] Avoiding spurious correlations would mean ensuring that no unrelated attributes are inadvertently correlated. For example, if in the training dataset most images of people cooking feature women instead of men, then the model might learn to recognize the action "cooking" based on whether there is a woman in the image.[128] In addition, determining the relevant subcategories within each group is a challenging task that AI developers are not necessarily best equipped to determine. For example, how many gender or ethnicity categories would be enough to ensure true diversity?

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 human image dataset or in models trained on it, it can be difficult

[125] This issue is known as "domain shift." *See, e.g.*, Daniel E. Ho, Emily Black, Maneesh Agrawala & Li Fei-Fei, *Evaluating Facial Recognition Technology: A Protocol for Performance Assessment in New Domains*, 98 DENV. L. REV. 753, 760–62 (2021).

[126] *See* DeVries et al., *supra* note 92, at 53.

[127] *See* Mitchell et al., *supra* note 1, at 220. It is difficult, however, to attain balance along so many dimensions — if you have three gender categories, three age groups, five ancestry groups, five scene types, three camera angles, and two illumination conditions, there will be 1,350 intersectional subgroups to balance data along.

[128] Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez & Kai-Wei Chang, *Men Also Like Shopping: Reducing Gender Bias Amplification Using Corpus-Level Constraints*, PROC. CONF. ON EMPIRICAL METHODS NAT. LANGUAGE PROCESSING 2979, 2979–80 (2017).

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 due to their training set not having 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 and does not fall into the trap of predatory inclusion.[129]

Moreover, uncovering bias in the first place can be difficult since existing publicly available datasets typically do not include people's self-reported demographics, so researchers or developers who want to ensure dataset diversity take measures to guess or estimate the demographics. Datasets with celebrities sometimes have web-scraped data on nationality.[130] When that information is not available, common methods include having annotators look at the photos and guess people's demographics,[131] using skin tone or other features as a proxy for race,[132] or using automated race classifiers.[133] While it would be ideal to collect self-reported demographics of the image subjects, collecting demographic data can present additional privacy concerns.[134] Without such data, however, even doing a preliminary check to see if the dataset is diverse or if the model performs well for different demographic groups is difficult.[135]

Given all these data collection challenges, the computer vision research community is divided on how important informed consent

[129] "Predatory inclusion refers to a process whereby members of a marginalized group are provided with access to a good, service, or opportunity from which they have historically been excluded but under conditions that jeopardize the benefits of access." Louise Seamster & Raphaël Charron-Chénier, *Predatory Inclusion and Education Debt: Rethinking the Racial Wealth Gap*, 4 SOC. CURRENTS 199, 199–200 (2017). The literature on this topic typically focuses on practices in the financial and educational domains. *Id.*

[130] Mei Wang, Weihong Deng, Jiani Hu, Xunqiang Tao & Yaohai Huang, *Racial Faces in the Wild: Reducing Racial Bias by Information Maximization Adaptation Network*, PROC. IEEE/CVF INT'L CONF. ON COMPUT. VISION 692, 693 (2019); Joo & Kärkkäinen, *supra* note 109.

[131] *See* Kärkkäinen & Joo, *supra* note 17.

[132] *See* Merler et al., *supra* note 9, at 2.

[133] *See* Wang et al., *supra* note 130, at 694.

[134] *See infra* Part VIII.

[135] Without such data, companies often rely on proxy variables. *See* Andrus et al., *supra* note 81, at 249. 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* Buolamwini & Gebru, *supra* note 6, at 1 (discussing shortcomings of using the Fitzpatrick skin tone scale as a proxy for race); Xiang, *supra* note 82, at 668 (discussing unintended consequences of using proxy variables for bias mitigation); Vidya Muthukumar et al., *Understanding Unequal Gender Classification Accuracy from Face Images*, ARXIV, Nov. 30, 2018, at 1, https://arxiv.org/abs/1812.00099 [https://perma.cc/SU4S-VUSU] (finding differences in performance are not due to skin tone).

should be for image datasets.[136] 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.[137] Even researchers who believed in the importance of informed consent stated they would still use datasets that do not have appropriate informed consent.[138] It was difficult for the researchers to see how they could conduct computer vision research and train accurate models otherwise.[139]

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 is used as evidence that developers do not care about such individuals. Even in situations where AI developers do care deeply about making their products work well for everyone, collecting sufficiently large and diverse datasets is very difficult and requires confronting many privacy and other ethical challenges.

VI. Privacy Laws

There are two separate considerations in privacy law that are relevant to the context of mitigating bias in computer vision systems: (1) the collection of biometric information and (2) the collection of sensitive attributes. The former is generally relevant for the development of any HCCV system but raises concerns in the context of collecting more diverse datasets, particularly from marginalized groups. The latter is important for bias detection and mitigation; it is difficult to evaluate dataset diversity or performance across demographic groups without demographic information.

Some of the most salient privacy laws in the first category are U.S. state laws like BIPA[140] and the California Consumer Privacy Act ("CCPA")[141] that regulate the processing of biometric information[142] and the EU's General Data Protection Regulation ("GDPR"), which restricts the processing of PII.[143] Biometric information can be seen as a

[136] Richard Van Noorden, *The Ethical Questions That Haunt Facial-Recognition Research*, 587 NATURE 354, 355 (Nov. 19, 2020), https://www.nature.com/articles/d41586-020-03187-3 [https://perma.cc/7LF9-8L4K].

[137] *Id.* at 357.

[138] *See id.* at 357–58.

[139] *See id.*

[140] 740 ILL. COMP. STAT. 14/15 (2008).

[141] CAL. CIV. CODE §§ 1798.100–.199 (2018) [hereinafter CCPA].

[142] Texas and Washington have also passed biometric information privacy laws. *See* Capture or Use of Biometric Identifiers Act, TEX. BUS. & COM. CODE § 503.001 (2009) [hereinafter CUBI]; WASH. REV. CODE § 19.375.020 (2017).

[143] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal

particularly sensitive subset of PII in this context. BIPA, for example, regulates the collection, storage, and use of biometric identifiers and biometric information.[144] Biometric identifiers include "scan[s] of hand or face geometry,"[145] which has been interpreted by courts to include both facial landmarks and facial templates.[146] CCPA's protections of biometric information more expansively include face images themselves (not just biometric information extracted from them), images of hands or palms, and gait patterns.[147]

While each jurisdiction's biometric information privacy laws differ slightly in scope, they all seek to restrict the collection, storage, and use of images or videos of faces or bodies (or landmarks and templates extracted from these images and videos) that *could* in turn be used to identify a person (actual use for identification is not required).[148] 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,[149] a right to request that the information be deleted,[150] a prohibition on denying goods or services for exercising privacy rights,[151] or a prohibition on the sale of or profit from the information.[152] The key protection this Article focuses on, however, is the requirement of informed consent to collect biometric information. While the type of notice and consent required varies under different laws,[153] some form of informed consent is the one constant across the various laws and, as discussed above in Part V, creates significant challenges in the development of HCCV.

The right to revoke consent under GDPR[154] also creates significant challenges for HCCV development. Even if a company goes through the steps of ensuring that they obtain informed consent and compensate individuals for their images, the fact that the data subjects might later

Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) [hereinafter GDPR].

[144] BIPA, *supra* note 42, § 15(b).

[145] *Id.* § 10.

[146] Yew & Xiang, *supra* note 41, at 1023.

[147] CAL. CIV. CODE § 1798.140(b).

[148] BIPA, *supra* note 42, § 15(b); GDPR, *supra* note 143, art. 4(1)–(2); CCPA, *supra* note 141, § 1798.100; CUBI, *supra* note 142, § 503.001(a)–(b); WASH. REV. CODE § 19.375.020 (2017).

[149] BIPA, *supra* note 42, § 15(b), (d); CUBI, *supra* note 142, § 503.001(b)–(c); GDPR, *supra* note 143, art. 12–15, 20; CCPA, *supra* note 141, §§ 1798.100, .110.

[150] GDPR, *supra* note 143, art. 17, 21; CCPA, *supra* note 141, § 1798.105.

[151] CCPA, *supra* note 141, § 1798.125(a)(1).

[152] BIPA, *supra* note 42, § 15(c). CCPA only requires disclosure to the customer if the information is sold and also provides the right for customers to opt out of such sales. CCPA, *supra* note 141, §§ 1798.115, .120.

[153] Some laws like BIPA require the data subject to provide written consent. *See* BIPA, *supra* note 42, § 15(b)(3). Others like CCPA only require notice with the right to access and delete any personal information collected. *See* CCPA, *supra* note 141, §§ 1798.100–1798.105.

[154] GDPR, *supra* note 143, art. 7(3).

revoke their consent means that companies must design systems to deal with such a possibility.[155] There is a lack of regulatory guidance, however, around the implications of such a revocation. While revocation does not affect the lawfulness of prior data processing,[156] what about future models derived from current models that used the data subject's image in development?[157] The need to enable individuals to revoke consent can also disincentivize companies from appropriately compensating individuals for their images. Given that GDPR requires consent to be freely given,[158] contractual provisions requiring the refund of the fee could be construed as undermining the extent to which the consent is completely voluntary.[159] This, however, can create a loophole by which individuals could receive payment for their images but revoke consent before the company is able to make significant use of those images.[160] In addition, enabling data subjects to revoke their consent ironically requires more data retention — if the images are completely stripped of any identifying information and an image subject then requests that their images be deleted, it will be difficult to determine which images feature them.[161]

Most of the U.S. privacy cases about image collection for HCCV center on BIPA. Over the past several years, BIPA's private right of action has made it a powerful tool for privacy advocates to challenge technology company data practices. Although state statutes like BIPA are narrow in their jurisdictional scope in theory, the difficulty in practice of determining whether images in a dataset are from Illinois residents has vastly expanded the influence of BIPA.[162] Part VII will

[155] *See generally* Eugenia Politou, Efthimios Alepis & Constantinos Patsakis, *Forgetting Personal Data and Revoking Consent Under the GDPR: Challenges and Proposed Solutions*, 4 J. CYBERSEC., no. 1, 2018, at 1.

[156] GDPR, *supra* note 143, art. 7(3) ("[T]he withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal.").

[157] As discussed *supra* note 40, it is common for general-purpose base models to be used as source models in the development of models for specific use cases.

[158] GDPR, *supra* note 143, art. 7(3).

[159] GDPR, *supra* note 143, recital 43 ("Consent is presumed not to be freely given if it does not allow separate consent to be given to different personal data processing operations despite it being appropriate in the individual case, or if the performance of a contract, including the provision of a service, is dependent on the consent despite such consent not being necessary for such performance.").

[160] A similar issue has been noted in the biomedical context. Politou et al., *supra* note 155, at 5 (citations omitted) ("[I]n the bio-banking field complete withdrawal could lead to the wastage of resources invested in bio-repositories . . . .").

[161] *See id.* ("[T]he same mechanisms [that] have been put in place to protect the privacy of data (like de-identification) may actually make it very difficult to trace and remove individual derived data in order to allow participants to withdraw completely their consent and be forgotten.").

[162] 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. *See, e.g.*, Jeffrey Neuburger, *Google App Disables Art-Selfie Biometric Comparison Tool in Illinois and Texas*, PROSKAUER NEW MEDIA & TECH. L. BLOG (Jan. 18, 2018), https://newmedialaw.proskauer.com/2018/01/18/google-app-disables-art-selfie-biometric-comparison-

feature a more in-depth discussion about the specific harms these laws seek to prevent and how courts have interpreted them.

In the second category — laws protecting sensitive attribute data — there is the GDPR, which regulates the processing of special categories of personal data like race.[163] There are also 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 in specific domains.[164] In practice, these restrictions have ironically erected significant barriers to both private and public sector entities attempting to audit their algorithmic systems for bias.[165] An interview study of algorithmic fairness practitioners found that companies across the AI industry, both small and large, overwhelmingly struggle to check their AI systems for bias, let alone take remedial measures to address bias.[166] Despite the growth in AI ethics, responsible AI, and algorithmic fairness teams in technology companies, these teams face practical challenges when attempting to convince their colleagues to collect sensitive attribute data to conduct bias assessments.[167] Legal and compliance teams often shut down efforts to collect, share, or use such data.[168] Considering existing privacy laws, this knee-jerk reaction is understandable, but it makes progress toward less-biased AI more challenging.

There is evidence that policymakers are increasingly cognizant of this challenge. The proposed EU AI Act creates a carve-out for processing sensitive data for bias monitoring, detection, and correction for

tool-in-illinois-and-texas [https://perma.cc/Y85N-BMLG]. Note, however, that Google did ask for consent from users of the app before processing their selfies. *Id.*

[163] GDPR, *supra* note 143, art. 9(1).

[164] 12 C.F.R. § 1002.5(b) (2012); *§ 1002.5 Rules Concerning Requests for Information*, CFPB, https://www.consumerfinance.gov/rules-policy/regulations/1002/5/#a-4-vi [https://perma.cc/4UMF-DSFQ].

[165] Scholars have shown how using anonymized smartphone-based mobility data to inform COVID-19 response strategies can perpetuate demographic biases. These biases can be difficult to detect due to the fact that such data is aggregated up from the individual level for privacy reasons. Amanda Coston, Neel Guha, Derek Ouyang, Lisa Lu, Alexandra Chouldechova & Daniel E. Ho, *Leveraging Administrative Data for Bias Audits: Assessing Disparate Coverage with Mobility Data for COVID-19 Policy*, PROC. ACM CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 173 (2021).

[166] *See* Andrus et al., *supra* note 81, at 257–58. Clavell et al. also discuss the challenge in practice of balancing data minimization under GDPR and bias audits. Gemma Galdon Clavell, Mariano Martín Zamorano, Carlos Castillo, Oliver Smith & Aleksandar Matic, *Auditing Algorithms: On Lessons Learned and the Risks of Data Minimization*, PROC. AAIM/ACM AI, ETHICS, & SOC'Y CONF. 265, 265, 269 (2020) https://dl.acm.org/doi/pdf/10.1145/3375627.3375852 [https://perma.cc/MN3Q-TZKT].

[167] Chloé Bakalar et al., *Fairness on the Ground: Applying Algorithmic Fairness Approaches to Production Systems*, META AI 1, 3 (2021), https://ai.facebook.com/research/publications/applying-algorithmic-fairness-approaches-to-production-systems [https://perma.cc/CY9R-RG9R].

[168] *See* Andrus et al., *supra* note 81, at 251–52.

high-risk AI systems.[169] 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.[170] However, there is less clarity on the U.S. side about how to balance privacy and bias mitigation. While there have been growing calls for audits of technology company algorithms,[171] there have not been changes to or guidance around privacy laws that would facilitate the collection of the personal information needed for conducting such 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 efforts toward less biased facial recognition systems.[172]

VII. Harms of Being Seen Versus 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 Part will focus on breaking down the specific harms of being "seen" and "mis-seen" 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.[173] In the seminal article *The Right to Privacy*, which

[169] *Proposed EU AI Act*, *supra* note 3, at 48 ("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.").

[170] *Guidance on AI and Data Protection*, INFO. COMM'R'S OFF., https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection [https://perma.cc/XY9Q-YSFX].

[171] Pauline T. Kim, *Auditing Algorithms for Discrimination*, 166 U. PENN. L. REV. ONLINE 189, 189–91 (2017), https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1212&context=penn_law_review_online [https://perma.cc/8KWE-CLJK]; James Guszcza, Iyad Rahwan, Will Bible, Manuel Cebrian & Vic Katyal, *Why We Need to Audit Algorithms*, HARV. BUS. REV. (Nov. 28, 2018), https://hbr.org/2018/11/why-we-need-to-audit-algorithms [https://perma.cc/PA4Z-L7T8]; Alex Engler, *Auditing Employment Algorithms for Discrimination*, BROOKINGS (Mar. 12, 2021), https://www.brookings.edu/research/auditing-employment-algorithms-for-discrimination [https://perma.cc/UJT7-Q39T].

[172] *See* Andrus et al., *supra* note 81, at 252–53.

[173] *See* Danielle Keats Citron & Daniel J. Solove, *Privacy Harms*, 102 B.U. L. REV. 793, 793–99 (2022).

is credited for essentially creating the U.S. common law right to privacy,[174] Samuel Warren and Louis Brandeis discussed privacy as "the right to be let alone."[175] The authors compared 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 defamed."[176] 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).[177] They justified 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."[178]

Modern U.S. 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"),[179] the Privacy Act of 1974 (applying to federal agencies),[180] sectoral federal statutory regulation, and state statutory regulation.[181] As discussed in Part VI, state biometric privacy laws like Illinois's BIPA and California's CCPA are most relevant to our discussion. These laws are notable for going beyond the sectoral nature of federal privacy laws, providing protections for biometric information or personal data regardless of the context of collection or use. While the right to privacy might largely 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.[182]

However, what are the specific harms that laws like BIPA protect against? Constitutional standing requires a concrete, particularized harm.[183] In *Spokeo, Inc. v. Robins*,[184] a case involving a "people search

[174] DANIEL J. SOLOVE & PAUL M. SCHWARTZ, INFORMATION PRIVACY LAW 10–11 (7th ed. 2021).

[175] Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 193 (1890).

[176] *Id.* at 205.

[177] *Id.* at 197–201.

[178] *Id.* at 198.

[179] 15 U.S.C. § 45(a)(1).

[180] 5 U.S.C. § 552a.

[181] SOLOVE & SCHWARTZ, *supra* note 174, at 812–13.

[182] 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: (1) informing the individual that the biometric identifier or information is being collected or stored, (2) informing the individual of the length of time of the collection, storage, or use, and (3) receiving written release from the individual. BIPA, *supra* note 42, § 15(b).

[183] Lujan v. Defs. of Wildlife, 504 U.S. 555, 560 (1992).

[184] 578 U.S. 330 (2016).

engine" where users would find detailed personal information about individuals, the Court found that "bare procedural violations" did not constitute a concrete injury sufficient for standing; the requirements of standing that an injury-in-fact be "concrete and particularized" had to be independently met.[185] After remand, the Ninth Circuit articulated 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."[186]

This test was later applied by the Ninth Circuit in *Patel v. Facebook, Inc.*,[187] a case (which Facebook ultimately settled for $650 million)[188] alleging that Facebook had violated BIPA in using facial recognition in its "Tag Suggestions" technology without obtaining appropriate informed consent from users.[189] In evaluating standing, 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."[190] The court thus stated that BIPA protects an individual's "concrete interests in privacy, not merely procedural rights."[191]

Key to the Ninth Circuit's finding was the idea that common law protects an individual's "control of information concerning his or her person,"[192] 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 their taxonomy of privacy harms, Citron and Solove discuss "lack of control" as one form of autonomy harm.[193] As they discuss, however, courts have been inconsistent in recognizing loss of control as a privacy harm.[194] Some courts have interpreted BIPA as only applying to the sharing of data with external parties — in *Rivera v. Google*,[195] the court

[185] *Id.* at 341.

[186] Robins v. Spokeo, Inc., 867 F.3d 1108, 1113 (9th Cir. 2017).

[187] 932 F.3d 1264, 1271 (9th Cir. 2019).

[188] Jennifer Bryant, *Facebook's $650M BIPA Settlement 'a Make-or-Break Moment,'* INT'L ASS'N PRIV. PROS. (Mar. 5, 2021), https://iapp.org/news/a/facebooks-650m-bipa-settlement-a-make-or-break-moment [https://perma.cc/E9PG-H8NN]; Nicholas Iovino, *Judge Approves Historic $650M Facebook Privacy Settlement*, COURTHOUSE NEWS SERV. (Feb. 26, 2021), https://www.courthousenews.com/judge-approves-historic-650m-facebook-privacy-settlement [https://perma.cc/8GTW-CYJU].

[189] *Patel*, 932 F.3d at 1268.

[190] *Id.* at 1273.

[191] *Id.* at 1265.

[192] *Id.* at 1273.

[193] Citron & Solove, s*upra* note 173, at 853–54.

[194] *Id.*

[195] 366 F. Supp. 3d 998 (N.D. Ill. 2018).

denied standing given that Google only stored (and had not shared) biometric data without consent.[196] While other courts have taken the opposite view, they did so by sidestepping the harm question and concluding that a violation of BIPA alone was sufficient for standing, even without actual injury or adverse effect.[197] In the HCCV cases litigated thus far in the United States, there has been no allegation of disclosure of information leading to mental harm, similar to the gossiping press that was decried by Warren and Brandeis as the impetus for a right to privacy. Instead, in evaluating Article III standing, the most concrete harms U.S. courts have identified are increased risks of identity theft and surveillance.

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*, the court expressed concern that the face templates collected by Facebook could be used to unlock cell phones.[198] It is unclear, however, that extracting face templates or landmarks from face images 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.[199] This can be done regardless of whether the images are also used to develop HCCV. If the developer is using publicly available images to develop the HCCV system, it is especially unclear that doing so would increase the risk of identity theft for the image subjects. This is not to say that having images publicized is not a harm in and of itself — indeed, the identity theft harm described above is a result of having images of yourself shared publicly — rather, it is important to distinguish the harm of having an image made public from the harm of having that public image used to train or evaluate HCCV.

While courts have appreciated the economic nature of identity theft harms, the most significant potential harm animating privacy fears around HCCV is the specter of mass surveillance. The harms in this context are related to safety concerns (e.g., a stalker finding your location) and chilling effects (e.g., self-censorship). In contexts where there is significant distrust of the government or disagreement about the appropriateness of the laws being enforced, being surveilled is also considered a societal harm. For example, there has been significant

[196] *Id.* at 1001, 1006.

[197] Rosenbach v. Six Flags Ent. Corp., 129 N.E.3d 1197, 1197–98 (Ill. 2019).

[198] *Patel*, 932 F.3d at 1273 ("Or a biometric face template could be used to unlock the face recognition lock on that individual's cell phone.").

[199] Lily Hay Newman, *Hackers Trick Facial-Recognition Logins with Photos from Facebook (What Else?)*, WIRED (Aug. 19, 2016, 8:00 AM), https://www.wired.com/2016/08/hackers-trick-facial-recognition-logins-photos-facebook-thanks-zuck [https://perma.cc/8CL6-F97D].

criticism of government efforts to surveil journalists[200] or opposition party members.[201] Moreover, one of the most controversial uses of mass surveillance is the Chinese government's tracking of the Uyghur ethnic minority group.[202]

Indeed, the potential for mass surveillance was a concrete harm the Ninth Circuit found to be compelling in *Patel*.[203] The court expressed concern that:

> Once a face template of an individual is created, Facebook can use it to identify that individual in any of the other hundreds of millions of photos uploaded to Facebook each day, as well as determine when the individual was present at a specific location . . . . It seems likely that a face-mapped individual could be identified from a surveillance photo taken on the streets or in an office building.[204]

When considering how to regulate the potential for HCCV to facilitate mass surveillance, several factors are important to consider. First, the harm of mass surveillance is tied specifically with the breadth of deployment of HCCV rather than the breadth of the data used to develop it. The harms discussed in *Patel* are specific to having one's image included in a reference set of images, against which new images are compared, so such harms are not directly relevant for the data used to train or test the model. The *Patel* case can thus be contrasted with *Flores v. Motorola Solutions, Inc.*,[205] in which mugshot images were used as a

[200] Basma Humadi, *Mass Surveillance Threatens Reporting that Relies on Confidential Sources*, REPS. COMM. FOR FREEDOM PRESS (Sept. 30, 2019), https://www.rcfp.org/nsa-mass-surveillance-against-journalist [https://perma.cc/TDK8-NGF9]; Emily Bell, Ethan Zuckerman, Jonathan Stray, Shelia Coronel & Michael Schudson, *Comment to Review Group on Intelligence and Communications Technologies Regarding the Effects of Mass Surveillance on the Practice of Journalism*, OFF. DIR. NAT'L INTEL. (Oct. 4, 2013), https://www.dni.gov/files/documents/RG/Effect%20of%20mass%20surveillance%20on%20journalism.pdf [https://perma.cc/ZKZ3-VWAR].

[201] *See, e.g.*, Özgün E. Topak, *The Making of a Totalitarian Surveillance Machine: Surveillance in Turkey Under AKP Rule*, 15 SURVEILLANCE & SOC'Y 535 (2017), https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/6614 [https://perma.cc/93ZL-VPP5]; Pinkaew Laungaramsri, *Mass Surveillance and the Militarization of Cyberspace in Post-Coup Thailand*, 9 AUSTRIAN J. SE. ASIAN STUD. 195 (2016), https://www.tde-journal.org/index.php/aseas/article/download/2648/2260 [https://perma.cc/C7P2-LFPJ].

[202] Drew Harwell & Eva Dou, *Huawei Tested AI Software That Could Recognize Uighur Minorities and Alert Police, Report Says*, WASH. POST (Dec. 8, 2020, 10:30 AM), https://www.washingtonpost.com/technology/2020/12/08/huawei-tested-ai-software-that-could-recognize-uighur-minorities-alert-police-report-says [https://perma.cc/756H-V9BX]; Mozur, *supra* note 106.

[203] Patel v. Facebook, Inc., 932 F.3d 1264 (9th Cir. 2019).

[204] *Id.* at 1273.

[205] No. 1:20-cv-01128, 2021 WL 232627 (N.D. Ill. Jan. 8, 2021).

reference set in a facial recognition tool sold to law enforcement.[206] In the latter case, the use of people's images directly implicated potential surveillance harms.

This distinction between the harms of images used in development versus deployment could also be used to explain the different judgments in *Vance v. Microsoft Corp.*[207] compared with *Vance v. Amazon.com Inc.*[208] In *Vance v. Microsoft*, in which plaintiffs alleged that Microsoft had violated BIPA by using IBM's "DiF" dataset "to improve the fairness and accuracy of its facial recognition products," the Western District of Washington dismissed the plaintiff's complaint under BIPA § 15(c) that Microsoft had "otherwise profit[ed] from" the biometric information of the plaintiffs.[209] The court concluded that since the plaintiffs did not allege that Microsoft "disseminated or shared access to biometric data through its products," there was no alleged violation of BIPA's profit provision.[210] The same court, however, did not dismiss the BIPA profit complaint in a similar case against Amazon[211] due to the possibility that Amazon used the images as part of the reference set for its facial recognition product.[212] Taken together, these cases suggest that the profit component of BIPA only applies in cases where the biometric information becomes part of the product itself.

This emphasis on whether the biometric information is part of the product is not very coherent in the context of HCCV given that the images used to train the model are arguably key to the product itself. Nonetheless, the way the court happened to apply this distinction in these cases suggests that the court had some intuition for the key distinction drawn throughout this Article between using individuals' data for development versus deployment. Using the images simply for training or testing during development, as in the *Microsoft* case, creates minimal privacy risks in comparison to including the images in a reference set used in deployment, as in the *Amazon* case. As will be discussed in Part VIII, making a distinction between development and deployment is one possible privacy law carve-out that could address the tensions discussed in this Article. Such a distinction would provide much greater

[206] *Id.* at *1.

[207] 534 F. Supp. 3d 1301 (W.D. Wash. 2021).

[208] 534 F. Supp. 3d 1314 (W.D. Wash. 2021).

[209] *Microsoft*, 534 F. Supp. 3d at 1309.

[210] *Id.*

[211] *Amazon*, 534 F. Supp. 3d at 1319.

[212] *Id.* at 1324. It is highly unlikely, however, that Amazon incorporated the Diversity in Faces dataset into a reference dataset that it sold alongside its Rekognition product. The "known" faces used to make a match must be identified in some way in order to be useful — it is not helpful to say that a new face is the same as another face scraped from Flickr unless there is more information about the face from Flickr (e.g., name, ID, or evidence that the person committed a crime). A reference set of random anonymous faces is not useful for recognition purposes.

clarity than the current guidance focusing on whether the data is part of the product.[213]

Another key consideration when evaluating the potential for mass surveillance harms is that not all forms of HCCV facilitate mass surveillance. Face, body, and object detection or classification 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 or body matches are shared. If FRT is used only locally on your phone to sort your photos, and the matches are not shared with the company or anyone else, then such technology arguably does not enable surveillance. These and other nuances will be discussed further in Part VIII, which proposes possible solutions for minimizing both the harms of being "seen" and "mis-seen."

Of course, data collected for one purpose can in theory be repurposed for another, so it is important to evaluate the extent to which the data can be used as a reference set. Images of random unidentified individuals would not be very useful for a reference set for face identification.[214] There needs to be some identifying information or meaning to the reference set in order for the model's inference to be meaningful — e.g., this is an image of Tom, or this is an image of someone you previously took pictures of, or this is an image of the suspect. The key inquiry then is how difficult it is, given the data available to the developer, to match an anonymous image with relevant identifying or contextual information. If the developer only has access to a public dataset of anonymous images, the risk of surveillance is relatively low compared to if the developer has extensive access to identifying information about individuals. The prevalence of reverse image search technology has dramatically lowered the barrier to identifying individuals in anonymous datasets,[215] but the harm still comes at the point of identification rather than inclusion in training or evaluation data. Thus, while courts have frequently expressed concerns about the potential for HCCV to

[213] In both cases, the court did not dismiss the plaintiff's unjust enrichment claim, so it is still possible that this line of cases stemming from the Diversity in Faces dataset could provide more clarity in the future about the extent to which "lack of control" holds legal weight as a privacy harm. Thus far, the caselaw does not provide much guidance around such harms.

[214] *See* BUOLAMWINI ET AL., *supra* note 48, at 6 ("Face identification software can only match the image of a face to a person for whom it already has some appearance information.").

[215] *See* Tina Sieber, *6 Fascinating Search Engines That Search for Faces*, MUO (May 7, 2022), https://www.makeuseof.com/tag/3-fascinating-search-engines-search-faces [https://perma.cc/V5C9-TD7T]; Kashmir Hill, *The Secretive Company That Might End Privacy as We Know It*, N.Y. TIMES (Nov. 2, 2021), https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html [https://perma.cc/9UXW-CTL8]; Kashmir Hill, *A Face Search Engine Anyone Can Use Is Alarmingly Accurate*, N.Y. TIMES (May 26, 2022), https://www.nytimes.com/2022/05/26/technology/pimeyes-facial-recognition-search.html [https://perma.cc/2GEJ-YBFB].

facilitate surveillance, it is important to consider these additional nuances to gauge the actual risk given a specific fact pattern.

Given the uncertainty around the extent to which identify theft and surveillance harms are relevant to the HCCV development process, how should we assess the residual risk? One approach would be to analogize the HCCV development process to "seeing" and learning from the world, as humans do. There is no legally cognizable harm associated with a human looking at non-explicit images that are readily available online. If the images have Creative Commons licenses that allow them to be used for commercial uses, the human might even be able to incorporate those images into a commercial product.[216] From this perspective, the fact that AI learns to "see" people by distilling image pixels into biometric information does not inherently change the harm equation. Under this view, it should be fine for developers to use publicly available images (with appropriate licenses) to develop HCCV.

On the other hand, given how controversial some HCCV technologies are, there is a strong argument that people should have some control over whether their images are being used to develop such technology. Much of the negative reporting about FRT datasets without informed consent has emphasized the harm associated with knowing that your personal data is contributing to technologies without your knowledge, particularly technologies you oppose.[217]

From a policy perspective, there are two ways to address this type of harm. One would be the current privacy regime, which emphasizes the importance of individual consent before one's images can be used for specific purposes. Another would be to increase the regulation of what is acceptable versus unacceptable HCCV such that individuals could feel more assured that the technology developed with their data is considered (at least by legislative consensus) to be societally acceptable. The former has the advantage of punting the question of acceptable use cases to the individual to decide but has the disadvantage of conflating control with signing an informed consent form. Depending on the context, such agreements can be difficult to understand or impossible to negotiate.[218] The latter has the advantage of ensuring that data

[216] For examples of Creative Commons licenses that allow commercial use, see *3.3 License Types*, CREATIVE COMMONS, https://certificates.creativecommons.org/cccertedu/chapter/3-3-license-types [https://perma.cc/3FN9-Q8TW].

[217] *See* Solon, *supra* note 115 ("[S]ome . . . photographers whose images were included in IBM's dataset were surprised and disconcerted . . . that their photographs had been annotated with details including facial geometry and skin tone and may be used to develop facial recognition algorithms."); Hill & Krolik, *supra* note 122 (describing individuals' negative reactions upon learning that they were featured in a large dataset used to develop FRT); Cade Metz & Kashmir Hill, *Here's a Way to Learn if Facial Recognition Systems Used Your Photos*, N.Y. TIMES (Feb. 1, 2021), https://www.nytimes.com/2021/01/31/technology/facial-recognition-photo-tool.html [https://perma.cc/4EU9-G8LE].

[218] Indeed, the fixation of modern privacy law on informed consent has been widely criticized in the literature. *See, e.g.*, Frederik Zuiderveen Borgesius, *Informed Consent: We Can*

will not be used for certain purposes but the disadvantage of relying heavily on regulators to carefully draw the line between acceptable and unacceptable use cases.

Thus, despite the growth in regulations around biometric information, there are significant ambiguities that remain around the specific harms envisioned by such regulations and how they would manifest in the context of developing HCCV. The most concrete harms courts have found compelling (identity theft and surveillance) depend on technical nuances that have not been addressed by courts and are distinct from the primary harm at hand — the use of data without consent to develop controversial technologies. This Section thus strived to clarify these harms and their relevance in this context. As courts consider such cases around the images used to develop HCCV and as policymakers consider how to regulate HCCV, such distinctions will be highly salient.

B. Harms of Being Mis-Seen

In this Section, I will focus on four specific harms of being "misseen": differences in service provision, security threats, allocative harms, and representational harms. All 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 or 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 such that individuals experience different levels of service.[219] For example, when Florida started using an FRT service to verify the identities of individuals applying for unemployment benefits, there were concerns about whether the technology performed well across demographics.[220] Individuals whose faces could not be verified by the FRT system had to resort to a video call with the service provider, with wait times of two to six hours.[221]

Do Better to Defend Privacy, 13 IEEE SEC. & PRIV. (2015), https://ieeexplore.ieee.org/abstract/document/7085952 [https://perma.cc/37EX-TAAC].

[219] *See* Michael A. Madaio, Luke Stark, Jennifer Wortman Vaughan & Hanna Wallach, *Microsoft AI Fairness Checklist*, PROC. CHI CONF. ON HUM. FACTORS COMPUTING SYS. 1, 1 (2020), http://www.jennwv.com/papers/checklists.pdf [https://perma.cc/DM72-2FCB] ("AI systems can fail to provide the same quality of service to some people as they do to others.").

[220] *See* Kylie McGivern, *Facial Recognition Meant to Stop Unemployment Fraud Is Blocking Legitimate Applicants*, ABC ACTION NEWS: WFTS TAMPA BAY (June 11, 2021, 4:20 PM), https://www.abcactionnews.com/news/local-news/i-team-investigates/facial-recognition-meant-to-stop-unemployment-fraud-is-blocking-legitimate-applicants [https://perma.cc/8Q7D-FNMU]; Ron Hurtibise, *Florida Continues to Require Identity Verification With ID.me*, GOVERNING (May 9, 2022), https://www.governing.com/security/florida-continues-to-require-identity-verification-with-id-me [https://perma.cc/45KU-ZRQL].

[221] *See* McGivern, *supra* note 220.

The second category of harm is security threats. In the face verification context, worse performance of the technology for certain groups could lead to unauthorized access. Especially in family contexts where individuals might look similar, such security issues could enable family members to unlock each other's phones, creating significant privacy risks.[222] Increasingly, face verification is also used for building security and for payments,[223] so significant discrepancies in the ability of such systems to work for different groups could lead to harms like home break-ins or unauthorized credit card use.

The third category is allocative harm. This is when an inaccuracy leads to a misallocation of a good or opportunity.[224] 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 addition, a study found that eye-tracking devices did not work as well for Asian participants as for other groups.[225] As such technology is increasingly used by educational institutions to determine whether students are paying attention and to detect cheating behavior,[226] such disparities in performance could lead to a higher risk of Asian students being incorrectly flagged for bad behavior.[227]

Finally, we have representational harms, when algorithmic systems represent certain groups in negative, offensive, or other problematic ways.[228] This type of harm is common for classification tasks since

[222] Karen Levy & Bruce Schneier, *Privacy Threats in Intimate Relationships*, 6 J. CYBERSEC., no. 1, 2020, at 1.

[223] *See* Sam Dean, *Forget Credit Cards — Now You Can Pay with Your Face. Creepy or Cool?*, L.A. TIMES (Aug. 14, 2020, 5:00 AM), https://www.latimes.com/business/technology/story/2020-08-14/facial-recognition-payment-technology [https://perma.cc/5RN8-XW5P]; s*ee, e.g.*, Caroline Spivack, *NYC Seeks to Curb Facial Recognition Technology in Homes and Businesses*, CURBED N.Y. (Oct. 8, 2019, 8:00 AM), https://ny.curbed.com/2019/10/8/20903468/nyc-facial-recognition-technology-homes-businesses [https://perma.cc/797E-RAEP].

[224] *See* BAROCAS ET AL., *supra* note 90, at 33; Madaio et al., *supra* note 219 ("AI systems can unfairly allocate opportunities, resources, or information."); Mohsen Abbasi, Sorelle A. Friedler, Carlos Scheidegger & Suresh Venkatasubramanian, *Fairness in Representation: Quantifying Stereotyping as a Representational Harm*, PROC. SIAM INT'L CONF. ON DATA MINING 801 (2019).

[225] Pieter Blignaut & Daniël Wium, *Eye-Tracking Data Quality as Affected by Ethnicity and Experimental Design*, 46 BEHAV. RSCH. METHODS 67, 75 (2014).

[226] Todd Feathers & Janus Rose, *Students Are Rebelling Against Eye-Tracking Exam Surveillance Tools*, VICE (Sept. 24, 2020, 9:00 AM), https://www.vice.com/en/article/n7wxvd/students-are-rebelling-against-eye-tracking-exam-surveillance-tools [https://perma.cc/RN7K-AL84].

[227] Face detection technology used for exam surveillance has also been found to perform more poorly for Black students. Todd Feathers, *Proctorio Is Using Racist Algorithms to Detect Faces*, VICE (Apr. 8, 2021, 12:49 PM), https://www.vice.com/en/article/g5gxg3/proctorio-is-using-racist-algorithms-to-detect-faces [https://perma.cc/EK2J-LZEB]. Notably, the bias evaluation for this model was done by an independent student researcher using FairFace, which is a dataset consisting of images from Flickr that lack informed consent. Lucy Satheesan, *Proctorio's Facial Recognition Is Racist.*, PROCTOR NINJA (Mar. 18, 2021), https://proctor.ninja/proctorios-facial-recognition-is-racist [https://perma.cc/R85X-6Z4K].

[228] *See* BAROCAS ET AL., *supra* note 90, at 21; Madaio et al., *supra* note 219 ("AI systems can reinforce existing societal stereotypes. AI systems can denigrate people by being actively

such tasks involve applying a label to an image. A famous computer vision example of representational harm was when Google Photos labeled an image of two Black individuals as an image of gorillas.[229] This harm can also occur with algorithms that determine which parts of images are the most relevant to focus on. In 2021, Twitter scrapped its image-cropping algorithm following revelations that this algorithm was more likely to crop out Black faces in favor of white faces.[230] Representational harms can also stem from existing biased trends in society. In the popular COCO dataset, images of technologically-oriented objects like keyboards and mice are more likely to feature men than women.[231] This can lead to HCCV models trained on COCO learning stereotyped representations. AI-powered image caption generators might consistently label images of women at a computer incorrectly as "a man sitting at a desk with a laptop computer," further perpetuating existing stereotypes.[232]

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, as discussed previously, Generative Adversarial Networks ("GANs")[233] trained to generate a synthetic image of an individual with long hair have been shown to also feminize the facial features of the individual.[234] By conflating long hair with feminine facial features, GANs perpetuate the stereotype that men have short hair and women have long hair. Similarly, an app designed to make faces look more attractive could be offensive if it does so by making skin look lighter, an artifact of learning cultural biases that consider lighter complexions to be more attractive.[235] Generative language models have also been

derogatory or offensive. AI systems can over- or underrepresent groups of people, or even treat them as if they don't exist.").

[229] Notably, this highly offensive harm seems to not have been directly solved yet. Tom Simonite, *When It Comes to Gorillas, Google Photos Remains Blind*, WIRED (Jan. 11, 2018, 7:00 AM), https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind [https://perma.cc/HM9T-R79Q].

[230] Rumman Chowdhury, *Sharing Learnings About Our Image Cropping Algorithm*, TWITTER (May 19, 2021), https://blog.twitter.com/engineering/en_us/topics/insights/2021/sharing-learnings-about-our-image-cropping-algorithm [https://perma.cc/5W9B-RU48].

[231] *See* Zhao et al., *supra* note 128, at 2985.

[232] *See* Wang et al., *supra* note 86, at 2.

[233] GANs are generative models that "create new data instances that resemble your training data . . . . [GANs] pair[] a generator, which learns to produce the target output, with a discriminator, which learns to distinguish true data from the output of the generator." *Introduction*, GOOGLE DEVS., https://developers.google.com/machine-learning/gan [https://perma.cc/GX6M-GURA] (last updated July 18, 2022).

[234] Balakrishnan et al., *supra* note 88, at 557.

[235] 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, and white woman. Prabhu et al., *supra* note 109.

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.[236]

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. As discussed previously, researchers at Facebook found that their tool had a harder time identifying objects in photos taken in developing countries.[237] Because their training data was disproportionately collected from developed countries, the model was more likely to recognize toothpaste on a sink in a higher-income household.[238] This is why, depending on the task, it is important not only to consider the demographic diversity of the people in the images but also factors like the geographic diversity of where the images are taken.

VIII. Approaches to Balancing Privacy and Bias Mitigation

While privacy laws generally protect against the harms of being "seen" without consent, the harms of being "mis-seen" are not directly mitigated. 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.[239] There are a few possible approaches for addressing this imbalance, as the Sections below will discuss.

One would be to create narrow carve-outs in the protections against being "seen" through privacy law. Another path would be to alleviate some of the concerns about being "seen" through participatory design, the use of trusted third parties to collect data, or privacy-preserving technological advances. A final approach would be to increase the protections against being "mis-seen." Note that this Part does not advocate for all these options equally but instead seeks to present a wide array of options and discuss the pros and cons of each.

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

[236] Elle Hunt, *Tay, Microsoft's AI Chatbot, Gets a Crash Course in Racism from Twitter*, GUARDIAN (Mar. 24, 2016), https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter?CMP=twt_a-technology_b-gdntech [https://perma.cc/DJ9Z-8GEA].

[237] *See* DeVries et al., *supra* note 92.

[238] *See id.* at 53.

[239] *See* Andrus et al., *supra* note 81, at 253.

for stronger privacy protections, and the specter of mass surveillance seems increasingly threatening with more and more deployment of HCCV technologies.[240] Indeed, some scholars have argued that we should instead be increasing privacy protections in the United States in order to prevent the ethical and legal risks associated with FRT.[241] Countries like China have recently increased privacy protections in the FRT context. The Supreme People's 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."[242] Nonetheless, given the tension presented in this Article between multiple ethical desiderata — privacy and fairness — and given efforts in the EU and UK to create privacy carve-outs for processing sensitive data in service of bias mitigation efforts,[243] it is worth contemplating what possible surgical changes could be made to existing privacy regimes to balance these desiderata.

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 consist of examples used to teach the model how to perform a specific task,[244] such as detection, recognition, or classification in a certain context. When an HCCV system learns *how* to identify the individual, the goal is not to identify that individual. Similarly, evaluation datasets are designed to assess the HCCV system's performance[245] rather than to make use of an identification. In contrast, when the HCCV system is deployed, the goal is to detect, recognize, or classify the individuals it encounters. For the task of recognizing a specific person, 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 are what directly enables mass surveillance.

Making this distinction between development and deployment would have the benefit of enabling HCCV developers to use large corpora of publicly available images and any other images they collect with appropriate licenses to train more accurate and less biased HCCV

[240] *See* Lindsey Barrett, *Ban Facial Recognition Technologies for Children — And for Everyone Else*, 26 B.U. J. SCI. & TECH. L. 223, 223 (2020); Sharon Nakar & Dov Greenbaum, *Now You See Me. Now You Still Do: Facial Recognition Technology and the Growing Lack of Privacy.*, 23 B.U. J. SCI. & TECH. L. 88, 89 (2017); Joel R. Reidenberg, *Privacy in Public*, 69 U. MIAMI. L. REV. 141, 141 (2014).

[241] *See* Nakar & Greenbaum, *supra* note 240.

[242] *Ruling by Top China Court Respects Privacy*, S. CHINA MORNING POST (Aug. 10, 2021, 11:30 PM), https://www.scmp.com/comment/opinion/article/3144579/ruling-top-china-court-respects-privacy [https://perma.cc/JGH4-N54E].

[243] *See supra* Part VI.

[244] *Training Data*, TECHOPEDIA, https://www.techopedia.com/definition/33181/training-data [https://perma.cc/R63D-ZATL] (last updated Feb. 17, 2022).

[245] Evaluation sets include both validation and test sets. *See id.*

systems. This could promote the creation of larger, more diverse publicly available datasets, leveling the playing field for smaller companies.

Of course, the drawback to this approach is that individuals would not be able to control what kinds of technologies their images are used to develop. Copyright would still apply, so the only images industry developers could use would be those that already have a license for commercial use, but many people would likely still feel uncomfortable if their images (even if publicly available with appropriate licensing) were used to develop HCCV.[246] Indeed, given that the copyright belongs to the photographer rather than the image subject, copyright might not provide any protection for many individuals.

In addition, to the extent images processed in deployment are used for further training of the model, the lines between development and deployment might blur. In these cases, the images should retain privacy protections to prevent such a carve-out from enabling additional surveillance use cases without appropriate informed consent.

Another possible approach would be to make the privacy laws around biometric information more domain-specific or sectoral. Indeed, federal privacy laws in the United States remain sectoral, protecting highly sensitive information in specific contexts, such as medical information.[247] The Health Information Portability and Accountability Act of 1996 ("HIPAA") is a key example.[248] One of the primary sources of imbalance between privacy and fairness considerations in HCCV development is the fact that antidiscrimination protections are highly sectoral, whereas the state biometric privacy protections are not.[249] 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 the rationale that biometric information is uniquely immutable, this innovation significantly expanded the scope of such laws.

Indeed, even the proposed EU AI Act, which some have criticized for being overly broad,[250] focuses specifically on prohibited use, high-

[246] *See supra* note 217 and accompanying text.

[247] SOLOVE & SCHWARTZ, *supra* note 174, at 974–75.

[248] Health Insurance Portability and Accountability Act (HIPAA) of 1996, Pub. L. No. 104–191.

[249] *See supra* Part VI.

[250] *DIGITALEUROPE's Initial Findings on the Proposed AI Act*, DIGITALEUROPE (Aug. 6, 2021), https://www.digitaleurope.org/wp/wp-content/uploads/2021/08/DIGITALEUROPEs-initial-findings-on-the-proposed-AI-Act.pdf [https://perma.cc/3TW3-5WWN]; Andrew McAfee, *EU Proposals to Regulate AI Are Only Going to Hinder Innovation*, FIN. TIMES (July 25, 2021), https://www.ft.com/content/a5970b6c-e731-45a7-b75b-721e90e32e1c [https://perma.cc/PQ4J-YYUJ]; *Feedback on the Artificial Intelligence Act*, CTR. DATA INNOVATION (2021), https://www2.datainnovation.org/2021-feedback-aia.pdf [https://perma.cc/A7ME-CSTA]; Dan Whitehead, *Hogan Lovells Responds to the European Commission's Consultation on the AI Regulation*, HOGAN LOVELLS (Aug. 10, 2021),

risk, or limited-risk cases of AI.[251] The regulation provides no requirements for other use cases. Similarly, privacy protections relevant to collecting and processing human images to develop HCCV could be limited to contexts like law enforcement, healthcare, finance, employment, education, and any other high-risk domains.

A likely critique of this approach, however, would be that many of the images used to develop HCCV are used to build general-purpose HCCV systems that can be tailored for a wide variety of different domains. For example, Amazon Rekognition offers a variety of pre-trained computer vision models, including face detection and analysis, that can be used for a wide variety of downstream uses.[252] As their guidelines discuss, intended use cases for their facial analysis models providing attributes like Smile, Pose, and Sharpness include selecting the "best profile picture" automatically in a social media application or anonymously estimating the gender and age of people at an event or retail store.[253] While they specifically disclaim that their emotion predictions do not necessarily reflect someone's internal emotional state,[254] their emotion recognition technology could in theory be used in controversial classroom monitoring, hiring, or suspicious behavior detection applications.

A more promising approach would be to make identifiability a salient factor when evaluating the collection or processing of biometric information. This would incentivize privacy-preserving techniques, such as blurring faces or manipulating them to be less recognizable, and efforts to silo data by storing identifying or sensitive information separately (with stricter access requirements) from other data to prevent matching or identification. Such a carve-out, however, will require significant guidance. As will be discussed in Section VIII.D, there are limitations to privacy-preserving techniques, such that the degree of identifiability that is relevant from a legal perspective will be a key question.

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 Act gestures in this

https://www.engage.hoganlovells.com/knowledgeservices/news/hogan-lovells-responds-to-the-european-commissions-consultation-on-the-ai-regulation [https://perma.cc/WTR8-KNXD].

[251] *See Proposed EU AI Act*, *supra* note 3, at 3.

[252] *Amazon Rekognition*, AMAZON WEB SERVS., https://aws.amazon.com/rekognition [https://perma.cc/LES7-DBM4].

[253] *Guidelines on Face Attributes*, AMAZON WEB SERVS., https://docs.aws.amazon.com/rekognition/latest/dg/guidance-face-attributes.html [https://perma.cc/XC9G-8S8G].

[254] *Id.*

direction with a carve-out for processing sensitive data to comply with other provisions in the regulation.[255] A similar approach could be used in the United States 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 would likely be highly controversial, this Section illuminates some possible carve-outs that would make privacy and fairness more evenly incentivized from a regulatory perspective. Overall, however, privacy carve-outs are the category of potential solutions with the most significant trade-offs, so pursuing the other possible solutions discussed below would be preferable.

B. Participatory Design

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 technology in its design to build greater trust between the data subjects and the data collectors. [256] In their piece discussing the parallels between data collection for AI and data collection for archives, Eun Seo Jo and Timnit Gebru emphasize the importance of establishing such community relationships and empowering communities to contribute to data collection efforts.[257]

This is an important approach to consider for bridging the gap between AI developers and communities affected by their development, but it faces many practical challenges to implementation. A key difference, however, between archives and datasets for AI 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,[258] contributing to an AI dataset is viewed with wariness. Many of the BIPA lawsuits against major U.S. technology companies came after people realized that their Flickr photos were being used in training datasets. An artist even created a platform for

[255] *See Proposed EU AI Act*, *supra* note 3, at 48.

[256] *See generally* Clay Spinuzzi, *The Methodology of Participatory Design*, 52 TECH. COMMC'N 163 (2005); MICHAEL J. MULLER & ALLISON DRUIN, PARTICIPATORY DESIGN: THE THIRD SPACE IN HUMAN-COMPUTER INTERACTION, *in* HUMAN COMPUTER INTERACTION HANDBOOK: FUNDAMENTALS, EVOLVING TECHNOLOGIES, AND EMERGING APPLICATIONS 1062 (Julie A. Jacko ed., 3d ed. 2012).

[257] Eun Seo Jo & Timnit Gebru, *Lessons from Archives: Strategies for Collecting Sociocultural Data in Machine Learning*, PROC. ACM CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 306, 306 (2020) https://arxiv.org/pdf/1912.10389.pdf [https://perma.cc/C4QA-HYQ8].

[258] *See generally Donating Your Materials to an Archive*, UNC UNIV. LIBRS., https://library.unc.edu/preservation/donating-your-materials [https://perma.cc/WBW6-JUCW]; *What Are Archives?*, KING'S COLL. CAMBRIDGE, https://www.kings.cam.ac.uk/archive-centre/introduction-to-archives/a/1 [https://perma.cc/FQ7G-BJLU].

people to check whether their images are included in the major publicly available datasets,[259] and journalists wrote of the creepiness of realizing their images were being used.[260]

The challenge for AI developers will thus 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 for everyone. In addition, community trust will likely be predicated on the images only being used to develop HCCV systems that the individuals believe will benefit their communities. However, much of the data used for training HCCV systems is used to train base models that can perform general tasks — such as object, face, or body detection, recognition, and verification — not specific to particular use cases.[261] 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 could be trained on many images from other communities. As a result, AI companies typically seek more global consent when using individuals' photos to develop any computer vision system.

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

C. Trusted Third-Party Data Collection

One method for addressing these trust issues is to shift the responsibility for data collection and storage from private companies to third-party actors (governmental or non-governmental) that might be more trusted for data collection. Michael Veale and Reuben Binns, for exam-

[259] EXPOSING.AI, https://exposing.ai [https://perma.cc/HJL9-M3VE]; Metz & Hill, *supra* note 217.

[260] *See* Hill & Krolik, *supra* note 122.

[261] *See, e.g.*, Conger, *supra* note 85; Shead, *supra* note 117. In these cases, billions of images were used to train general models to do face verification and object recognition. In general, the reason transfer learning is so popular is that large numbers of images and amounts of computing are required to train deep learning models, so it is standard to adapt a large model trained for a general purpose for more specific tasks. Jason Brownlee, *A Gentle Introduction to Transfer Learning for Deep Learning*, MACH. LEARNING MASTERY (Dec. 20, 2017), https://machinelearningmastery.com/transfer-learning-for-deep-learning [https://perma.cc/4END-V99U].

ple, have proposed this approach as a way to handle the privacy concerns around processing sensitive attribute data for bias mitigation.[262] This would have the advantage of creating large image datasets using the pooled resources of companies, research institutions, and/or government entities, alleviating some of the challenges facing developers that do not have a preexisting pipeline for images. In addition, if 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. If the data consortium succeeds in being a trustworthy entity, then more people will likely be willing to contribute data to the entity in comparison to selling their data to companies with weaker ethical governance guarantees. The presence of such a trusted data consortium would also raise the ethical standards for data collection — even when companies are collecting their own data, their practices could be compared to those of the consortium.

Creating a third-party trusted data consortium to collect HCCV data would have to go beyond the proposal of Veale and Binns, however, given the need not only to manage the sensitive attribute data used to audit an AI system but also the fundamental building blocks of the HCCV system itself. The complications and challenges of ethical data collection discussed above persist. The third-party entity will have to struggle with questions of how to collect a globally representative dataset with adequate informed consent and sufficiently candid and diverse images. While this solution directly tackles the problem of trust, it does not necessarily solve the other challenges.

Nonetheless, a third-party entity would arguably be better equipped to address some of these challenges. Economies of scale are quite useful for data collection generally, but especially for ethical and legally compliant sensitive data collection, given the high overhead costs. For example, designing informed consent agreements that are compliant with all jurisdictions' privacy regulations is quite challenging, but the cost of designing such agreements is the same regardless of whether thousands or millions of images are collected. In addition, setting up systems for data subjects to monitor how their data is being used and revoke their consent at will is quite difficult for small companies or academics, but could be more easily handled by a dedicated data consortium. More generally, there are high fixed costs to establishing a crowdsource platform where individuals around the world can upload

[262] Michael Veale & Reuben Binns, *Fairer Machine Learning in the Real World: Mitigating Discrimination Without Collecting Sensitive Data*, 4 BIG DATA & SOC'Y, no. 1, Dec 2017, at 1, 1.

their images and be compensated fairly. Even when large companies have tried to create such platforms, their initiatives have not been successful.[263]

It is difficult, however, 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 police mugshots and images of exploited children (among other marginalized populations) as the basis for their test dataset to evaluate commercial facial recognition systems,[264] 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. A new entity might have to be created to take on this responsibility.[265]

Such a solution will likely take years to develop, so it alone will not be a panacea. Moreover, even if such a trusted data consortium exists in the future, companies will still need to collect some of their own data to tailor their models to specific tasks.[266] Nonetheless, given the long-term benefits, creating trusted third parties to collect data for HCCV development is a promising solution to pursue alongside other approaches.

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 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

[263] Microsoft created a platform called Trove in 2020 to enable the responsible collection of images from crowd workers. Over 60,000 images were collected. Microsoft shut down the Trove project on November 16, 2021. *Microsoft Trove*, MICROSOFT, https://www.microsoft.com/en-us/ai/trove-images-for-machine-learning [https://perma.cc/U4J5-DTUV].

[264] *See* GROTHER ET AL., PART 1, *supra* note 105; *Chexia Face Recognition*, *supra* note 104.

[265] It is also possible that existing non-profit organizations could host an initiative like this. For example, Mozilla currently has an initiative for collecting voice data. *Common Voice*, MOZILLA, https://commonvoice.mozilla.org/en [https://perma.cc/38YJ-E95D].

[266] The large image dataset provided by a trusted third party is useful for training a base model that is able to do common tasks like identification or verification, but 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 companies will still need to address the ethical dataset collection challenges discussed in this Article.

formal privacy guarantee that the original image cannot be reverse-engineered.[267] While completely cutting out an individual's face can dramatically reduce the possibility of future identification, such images are only useful for developing non-face-related HCCV. Moreover, such techniques 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.[268] The process of face blurring itself requires processing biometric information, creating a catch-22. More generally, there is always a trade-off between the level of privacy attained through such techniques and the utility of the data.[269] This is not to say that privacy-preserving techniques are not an important part of HCCV systems, but rather that they alone cannot solve the problems discussed in this Article.

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 or settings, thus augmenting a given training dataset. Typically, 3D scans are performed for a set of individuals and then generative models are applied to expand this data into many more combinations[270] by modifying specific features of an individual (e.g., skin tone, hair length, or perceived gender)[271] or "hallucinating" new people.[272] Such generative models, however, need to be trained on large numbers of human images,[273] thus undermining the extent to which this

[267] William L. Croft, Jörg-Rüdiger Sack & Wei Shi, *Differentially Private Obfuscation of Facial Images*, 3 PROC. INT'L CROSS-DOMAIN CONF. FOR MACH. LEARNING & KNOWLEDGE EXTRACTION 229 (2019), https://link.springer.com/chapter/10.1007%2F978-3-030-29726-8_15 [https://perma.cc/45US-8RXT]. 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. *See Differential Privacy*, HARV. UNIV. PRIV. TOOLS PROJECT, https://privacytools.seas.harvard.edu/differential-privacy [https://perma.cc/K6PB-HPR4]; Alexandra Wood et al., *Differential Privacy: A Primer for a Non-Technical Audience*, 21 VAND. J. ENT. & TECH. L. 209, 232 (2018). In practice, differential privacy is achieved by adding random noise from a carefully chosen distribution to the data. Cynthia Dwork, Frank McSherry, Kobbi Nissim & Adam Smith, *Calibrating Noise to Sensitivity in Private Data Analysis*, 3 THEORY OF CRYPTOGRAPHY CONF. PROC. 265, 265 (2006), https://link.springer.com/chapter/10.1007/11681878_14 [https://perma.cc/Y2UN-FVQJ].

[268] Yew & Xiang, *supra* note 41, at 1022.

[269] *See, e.g.*, Croft et al., *supra* note 267.

[270] Karen Hao, *These Creepy Fake Humans Herald a New Age in AI*, MIT TECH. REV. (June 11, 2021), https://www.technologyreview.com/2021/06/11/1026135/ai-synthetic-data [https://perma.cc/2ADD-PCTQ].

[271] *See, e.g.*, Balakrishnan et al., *supra* note 88.

[272] *See, e.g.*, Blaž Meden, Žiga Emeršič, Vitomir Štruc & Peter Peer, *k-Same-Net: k-Anonymity with Generative Deep Neural Networks for Face Deidentification*, 20 ENTROPY, no. 1, Jan. 2018, at 1, https://ieeexplore.ieee.org/document/7985521 [https://perma.cc/TMG4-2PHT].

[273] *See, e.g.*, Tero Karras, Samuli Laine & Timo Aila, *A Style-Based Generator Architecture for Generative Adversarial Networks*, 43 IEEE TRANSACTIONS ON PATTERN

approach can completely resolve the informed consent barrier. While some companies have claimed to use "zero data,"[274] some images of real people are typically used somewhere in the pipeline of creating images of synthetic individuals.[275]

Other issues with synthetic data include realism and potential biases. Synthetic humans still visually appear distinct from real humans,[276] which can undermine the extent to which they can completely replace real human images in the immediate term. This concern will likely be mitigated over time with advances in this type of technology, propelled by the demand for ever-more realistic-looking images. The second issue is more complicated to address. To the extent that only a small number of people are scanned to form the basis of the synthetic images, the synthetic images still might not accurately reflect the wide diversity of humanity.[277] Moreover, the people creating these images will inevitably have preconceptions of what are relevant types of people and contexts to feature.[278] 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 can 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 sidestep 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 en-

ANALYSIS AND MACH. INTEL. 4217, 4226 (2021) ("[W]e thus increase the training time from 12M to 25M images.").

[274] Some start-ups have claimed to be "zero data" by having artists create 3D models instead of relying on real ones. *See, e.g.*, Sage Lazzaro, *AI Experts Refute Cvedia's Claim Its Synthetic Data Eliminates Bias*, VENTUREBEAT (July 6, 2021, 2:20 PM), https://venturebeat.com/2021/07/06/ai-experts-refute-cvedias-claim-its-synthetic-data-eliminates-bias [https://perma.cc/9HT5-4BPF].

[275] Approaches to create realistic humans typically do involve at least some images of real people at the beginning to create the 3D models. *See, e.g.*, Hao, *supra* note 270; *Synthetic Data Case Studies: It Just Works*, SYNTHESIS AI (June 17, 2021), https://synthesis.ai/2021/06/17/synthetic-data-case-studies-it-just-works [https://perma.cc/HPZ8-6NHB].

[276] *See* Hao, *supra* note 270. For images of synthetic humans, see DATAGEN, https://datagen.tech [https://perma.cc/7VXH-C87S], and SURREAL DATASET, https://www.di.ens.fr/willow/research/surreal/data [https://perma.cc/76RM-78CL], and *PeopleSansPeople: A Synthetic Data Generator for Human-Centric Computer Vision*, UNITY TECHS., https://unity-technologies.github.io/PeopleSansPeople [https://perma.cc/5J5K-ZHU5].

[277] For example, studies have shown the tendency of GANs to amplify masculine facial features and fair skin tones. Niharika Jain, Alberto Olmo, Sailik Sengupta, Lydia Manikonda & Subbarao Kambhampati, *Imperfect ImaGANation: Implications of GANs Exacerbating Biases on Facial Data Augmentation and Snapchat Face Lenses*, 304 A.I., Mar. 2022, at 1, 1.

[278] *See* Lazzaro, *supra* note 274.

ables data across different parties to be used for training a model, without directly sharing that data between parties.[279] 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 have their images used for training HCCV but are uncomfortable with directly sharing their images with the entity in question. For example, someone might be comfortable with their photos being used for training an Apple photo-sorting 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 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 of better-performing, less-biased HCCV, federated learning can ease some concerns around sharing their data. In recent lawsuits around HCCV, however, courts have considered the distinction between whether images are stored on the user's edge device versus the company's cloud to be irrelevant for reducing a company's liability (the company was still considered to have control over the data).[280] While these cases are either still in progress or were settled before a decision was made, they suggest that using techniques like federated learning might not completely resolve the privacy challenges to creating less biased HCCV.

Thus, the tension between privacy and fairness in HCCV data collection might be reduced 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 about how to generate large numbers of diverse, realistic images without substantial bias.

E. Right Against Being Mis-Seen

The final approach to balancing the desire not to be "seen" or "misseen" would be to increase the protections against being "mis-seen." As discussed above in Section VII.B, 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

[279] Brendan McMahan & Daniel Ramage, *Federated Learning: Collaborative Machine Learning Without Centralized Training Data*, GOOGLE A.I. BLOG (Apr. 6, 2017), https://ai.googleblog.com/2017/04/federated-learning-collaborative.html [https://perma.cc/Z4NA-9TUJ].

[280] *See, e.g.*, Hazlitt v. Apple Inc., 543 F. Supp. 3d 643, 652–54 (S.D. Ill. 2021) ("Because BIPA merely requires that an entity 'possess' biometric data to be liable, not that the entity 'store' data on its own servers, Apple's argument that it is exempt from BIPA because its biometric data is located in databases on user devices that Apple alone controls is meritless.").

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. This Section will explore what a right *not* to be "mis-seen" might look like.

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 the existing product liability doctrine that would render it unable to provide sufficient protections.

First, strict product liability protects primarily against personal injury or property damage.[281] While robots and autonomous vehicles might be subject to such liability, many HCCV systems are primarily deployed in digital contexts where the potential for personal injury or property damage is minimal (e.g., verifying someone's identity, sorting photos, monitoring people, or providing entertainment on social media). Though 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, someone experiencing physical harm is still typically necessary for strict product liability.[282] Negligence law might apply some protection against a broader range of harms from HCCV systems, but scholars have noted many challenges to successfully applying negligence liability in the context of AI, including the difficulties of foreseeing AI errors.[283]

A second limitation is that product liability law would not help plaintiffs who experienced algorithmic bias.[284] 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.[285] 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. If the HCCV developer made false claims about the system being unbiased, the plaintiff might be able

[281] *Understanding the Interplay Between Strict Liability and Product Liability*, LEXISNEXIS (Jan. 6, 2021), https://www.lexisnexis.com/community/insights/legal/b/thought-leadership/posts/understanding-the-interplay-between-strict-liability-and-products-liability [https://perma.cc/9F9D-E7Y9].

[282] *See* Linda Trummer-Napolitano, *Emotional Distress in Products Liability: Distinguishing Users from Bystanders*, 50 FORDHAM L. REV. 291, 291–95 (1981); Jane B. Silverman, *Recovery for Emotional Distress in Strict Products Liability*, 61 CHI.-KENT L. REV. 545, 545–48 (1985).

[283] Andrew D. Selbst, *Negligence and AI's Human Users*, 100 B.U. L. REV. 1315, 1315 (2020).

[284] Selbst discusses the challenges of applying negligence law to contexts where there are unevenly distributed harms. *Id.* at 1354.

[285] *Id.* at 1315–16.

to succeed on a claim of false advertising,[286] but there is no such thing as a completely unbiased AI model, so such litigation would likely simply lead companies to avoid making such outlandish claims.

A third limitation is the lack of robust standards for performance in the AI industry. Industry standards are often relied upon in product liability law to evaluate whether a company has been negligent.[287] While NIST has created a Facial Recognition Vendor Test for companies to benchmark their facial recognition technologies,[288] there is no industry-wide consensus on a single benchmark for performance or what levels of performance are sufficient, particularly for demographic subgroups. Moreover, the need to tailor AI systems to specific deployment contexts suggests that any blanket benchmark or performance standard would be misleading.[289] 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 grainier and lower quality and rarely feature a clear frontal image of the suspect looking into the camera.[290]

While consumer expectations are also often used as a benchmark for reasonableness, HCCV is a relatively new and rapidly evolving technology, meaning expectations are particularly unstable.[291] This lack of clear consumer expectations has also made it easy for AI technologies to proliferate while providing minimal representations and warranties to consumers.[292] AI companies often avoid providing any details about how their technologies are developed or how well they perform on any standardized tests.[293]

Finally, there are many reasonable justifications for why companies do not do more to ensure their computer vision products are not

[286] For example, on the federal level, the Lanham Act prohibits "misrepresent[ing] the nature, characteristics, [or] qualities [of] . . . goods, services, or commercial activities." 15 U.S.C. § 1125(a)(1)(B).

[287] *See* Selbst, *supra* note 283, at 1353.

[288] *Face Recognition Vendor Test*, NAT'L INST. OF STANDARDS & TECH., https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt [https://perma.cc/SV7Y-S58U].

[289] Alicia Solow-Niederman, YooJung Choi & Guy Van den Broeck, *The Institutional Life of Algorithmic Risk Assessment*, 34 BERKELEY TECH. L.J. 705, 707 (2019). Notably, popular datasets like Labeled Faces in the Wild specifically warn that they should not be used for concluding whether an algorithm would be suitable for commercial purposes, citing a lack of diversity along age, gender, ethnicity, lighting conditions, poses, occlusions, and photo resolution. LABELED FACES IN THE WILD, http://vis-www.cs.umass.edu/lfw [https://perma.cc/QX28-GNWJ].

[290] *See, e.g.*, Pei Li, Patrick J. Flynn, Loreto Prieto & Domingo Mery, *Face Recognition in Low Quality Images: A Survey*, 1 ACM COMPUT. SURV., no. 1, Apr. 2019, at 1, 1–2.

[291] Selbst, *supra* note 283, at 1325.

[292] *See* Stefano Puntoni, Rebecca Walker Reczek, Markus Giesler & Simona Botti, *Consumers and Artificial Intelligence: An Experiential Perspective*, 85 J. MKTG. 131, 132–35 (2020).

[293] Selbst, *supra* note 283, at 1322 (discussing the importance of secrecy in AI development).

highly biased, making it difficult to pursue a negligence case. 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 to improve such performance.[294] Current industry practices around preventing algorithmic bias are often minimal due to the lack of incentives (and strong disincentives) to address this issue.[295]

Thus, new regulations at the state or federal level 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 antidiscrimination right for systems to not have a significantly disproportionate performance for one's 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 to develop specific HCCV regulations. Transparency obligations could further enable individuals to challenge the use of HCCV systems with poor performance.

The antidiscrimination 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 designed for them. Given that HCCV models lack intentionality and algorithmic bias is typically unintentional,[296] the protection would be against disparate impact, a form of unintentional discrimination whereby facially neutral practices lead to disproportionate adverse effects on particular subgroups.[297] While most antidiscrimination laws apply to specific domains, like employment, finance, or education, this protection would apply to a category of technology, HCCV. The justification for singling out HCCV for additional antidiscrimination protections would be that (1) bias mitigation for HCCV is particularly important but difficult[298] and (2) the increasing pervasiveness of HCCV in everyday contexts makes the lack of protections

[294] *See supra* Part VI.

[295] *See* Andrus et al., *supra* note 81, at 258.

[296] 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. *See* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 674–84 (2016); Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L.J. 1043, 1088–91 (2019).

[297] *See Title VI Legal Manual § VII*, U.S. DEP'T OF JUST. (2021), https://www.justice.gov/crt/fcs/T6Manual7 [https://perma.cc/55CV-5A73].

[298] *See supra* Part V.

against bias in HCCV particularly pernicious, even in low-stakes contexts.[299] While the domain-specificity of many antidiscrimination laws is motivated by the high-stakes nature of those contexts, there are also antidiscrimination laws like Title II and Title III of the Civil Rights Act of 1964 that protect individuals in lower-stakes but commonplace contexts like public accommodation.[300] 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.[301]

Having an antidiscrimination 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 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.[302] 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.[303]

[299] *See supra* Part III.

[300] 42 U.S.C. §§ 2000a–2000b.

[301] 42 U.S.C. §§ 12101–12213.

[302] A requirement for such disparate impact assessments was notably missing in the proposed EU AI Act. Mark MacCarthy & Kenneth Propp, *Machines Learn That Brussels Writes the Rules: The EU's New AI Regulation*, LAWFARE (Apr. 28, 2021, 10:51 AM), https://www.lawfareblog.com/machines-learn-brussels-writes-rules-eus-new-ai-regulation [https://perma.cc/A8Q9-DGU9].

[303] For a more in-depth discussion about relevant model-related disclosures for transparency purposes, see Mitchell et al., *supra* note 1, and Clavell et al., *supra* note 166 (discussing the importance of early documentation for enabling audits).

IX. Conclusion

Few technologies are currently as controversial as HCCV, prompting the passage of a flurry of privacy laws and moratoriums 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 or classifications. Not all HCCV enables mass surveillance, however, and the development of fair and accurate 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. 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 relationships of trust with the communities they seek to collect data from. There is no silver bullet for enabling more ethical HCCV systems that balance all 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 support for trusted third-party data collection initiatives, greater legal protections against being "mis-seen," and more clarity around acceptable uses of biometric and sensitive information for bias mitigation. Ultimately, we must balance the desire not to be "seen" with the desire not to be invisible.