Via Electronic Submission

June 30, 2022

Mason Clutter, Acting Executive Director
Privacy and Civil Liberties Oversight Board
800 North Capitol Street NW, Suite 565
Washington, DC 20002

Re: Public Comment Regarding Public Forum on Domestic Terrorism, Notice PCLOB-2022-01

The undersigned are affiliated with the Institute for Technology, Law & Policy (ITLP) and Article 19. ITLP is a collaboration between the UCLA School of Law and the UCLA Samueli School of Engineering whose mission is to foster research and analysis to ensure that new technologies are developed, implemented, and regulated in socially beneficial, equitable, and accountable ways. Article 19 is an international human rights organization focused on ensuring that everyone can freely express themselves and actively engage in public life without fear of discrimination.

We are grateful for the opportunity to submit a public comment and invite the Board to take the concerns below into consideration. While there are a wide range of privacy and civil liberties impacts related to U.S. efforts to counter domestic terrorism, this comment primarily focuses on those that intersect with modern information and communications technology. Our overarching recommendation is that there is a pressing need to carefully consider the adverse impacts of technologically-based initiatives aimed at countering terrorism, particularly those involving machine learning and facial recognition, as the biases within these systems tend to have an outsized effect on marginalized and minority groups. The same is true for pervasive surveillance measures, which may chill freedom of expression, particularly among vulnerable communities. Private sector enforcement, and the increasing role of private institutions or public-private partnerships in efforts to combat terrorism, leads to its own challenges in developing an appropriate model for protecting constitutional rights. While none of this is an argument for rolling back the clock on the use of new technologies altogether, it does reinforce the need for a robust and public auditing procedure and assessments to consider all downstream impacts of these policies.

1. Introduction

Whereas in previous administrations efforts were focused on countering violent extremism related to jihadist groups,[1] the increasing growth of violent White extremist movements across the United States has been identified by the current administration as an elevated and evolving threat.[2] Domestic terrorism is often rooted in violent ideologies, racism, ethnic bigotry, and anti-government sentiments, in most cases targeting marginalized and racialized communities.[3] Malicious actors have used technology to coordinate

[1] Julia Edwards Ainsley, Dustin Volz & Kristina Cooke, *Exclusive: Trump to Focus Counter-Extremism Program Solely on Islam - Sources*, REUTERS, (Feb. 3, 2017), https://www.reuters.com/article/us-usa-trump-extremists-program-exclusiv-idUSKBN15G5VO; The White House, *supra* note 1.

[2] The White House, *FACT SHEET: The White House Summit on Countering Violent Extremism*, OFFICE OF THE PRESS SECRETARY (Feb. 18, 2015), https://www.whitehouse.gov/wp-content/uploads/2021/06/National-Strategy-for-Countering-Domestic-Terrorism.pdf; Betsy Woodruff Swan, DHS Draft Document: White Supremacists Are Greatest Terror Threat, POLITICO, (Sept. 4, 2020), https://www.politico.com/news/2020/09/04/white-supremacists-terror-threat-dhs-409236.

[3] See e.g., Simon Romero & Nicholas Bogel-Burroughs, *El Paso Shooting: Massacre That Killed 20 Being Investigated as Domestic Terrorism*, N. Y. TIMES, (Aug. 4, 2019), https://www.nytimes.com/2019/08/04/us/el-paso-

terrorist acts, recruit, and spread extremist ideologies and propaganda, and to publicize and promote terrorist acts online.[4]

National security agencies have increasingly relied on predictive and identification technologies to detect domestic terrorist threats; private companies have been pressured to share user data, restrict, and remove content and regulate access to their services; and public-private partnerships have been promoted as instrumental in the fight against terrorism. However, there are a range of risks with these technologies and how they are governed which pose serious concerns for civil and political liberties as well as human rights, particularly with respect to marginalized and racialized minorities.

Inherent systemic, institutional, and technical biases present in the development and use of machine learning-based technology to counter domestic terrorism disproportionately impact racialized and marginalized communities and divert resources from equally concerning issues. The use of mass surveillance technology leads to chilling effects against the rights to freedom of expression, association, and other fundamental human rights, both online and offline. Efforts to detect and remove terrorist content online also hinder expressive rights of disadvantaged groups. Content moderation systems' inability to understand context and linguistic nuances, and the vague and discriminatory definitions used by private companies, have resulted in over-removal and over-profiling of marginalized and minority groups, as well as in the erroneous removal of journalist content, undermining efforts to report and denounce illegalities. Additionally, governments' direct and indirect involvement in content moderation practices further exacerbates risks posed to individuals' privacy, freedom of expression, and association. Public-private partnerships in national security efforts, where private companies have been turned into a new type of enforcers of government policy, also create novel constitutional challenges, especially considering the opacity that characterizes such agreements and the lack of effective public oversight.

2. The technology used in the fight against terrorism

This section examines how human rights – comprising civil rights, political liberties, and fundamental rights – are affected by technologies that are used in domestic efforts to combat terrorism and violent extremism. For example, national security initiatives have relied on automated surveillance to detect illegal or suspicious activity.[5] Machine learning and artificial intelligence have been used to develop tools such as natural language processing, facial recognition softwares, and algorithmic screening systems**,** which undergird the surveillance apparatus and the ICTs that define the modern public sphere where civil rights and political liberties are most consequential.[6] To meet the preventive and reactive goals of anti-terrorism

shooting-updates.html; Joe Heim et al., *One dead as car strikes crowds amid protests of white nationalist gathering in Charlottesville; two police die in helicopter crash*, WASH. POST, (Aug. 13, 2017), https://www.washingtonpost.com/local/fights-in-advance-of-saturday-protest-in-charlottesville/2017/08/12/155fb636-7f13-11e7-83c7-5bd5460f0d7e_story.html ; https://www.nytimes.com/2018/07/17/us/mosque-arson-guilty-verdict.html.

[4] Robert Graham, *How Terrorists Use Encryption*, 9:6 CTC SENTINEL 20 (2016); Mia Sato, *How the Buffalo shooting livestream went viral*, THE VERGE, (May 17, 2022 2:46 PM), https://www.theverge.com/2022/5/17/23100579/buffalo-shooting-twitch-livestream-viral-content-moderation; Jan Christoffer Andersen & Sveinung Sandberg, *Islamic State Propaganda: Between Social Movement Framing and Subcultural Provocation*, 32:7 TERRORISM AND POLITICAL VIOLENCE 1506 (2020).

[5] Didier Bigo et al., *Mass Surveillance of Personal Data by EU Member States and Its Compatibility with EU Law*, (Liberty and Security in Europe Papers No. 61, 2013).

[6] ARTICLE 19, *When bodies become data: biometric technologies and freedom of expression*, (Apr. 2021), https://www.article19.org/wp-content/uploads/2021/04/A19-Biometric-technologies-and-FoE-Policy-2021.pdf

measures – where rapidness tends to outweigh accuracy – speed and efficiency in data collection and processing through machine learning have become essential.[7] To understand the challenges posed by the use of technology in anti-terrorism, it is first necessary to identify what kind of technology is mostly used, and what are its limitations.

3. Machine Learning & A.I.

Machine learning systems collect, process, and store information (data) that can be used to analyze, predict, and make assessments about an individual or a group to which the individual belongs. Machine learning systems are developed and deployed in three main steps. First, developers identify the main goal of the system, e.g., identifying texts glorifying terrorist ideologies. Second, a database is selected on which to train the system, e.g., a database with examples of texts. During the training phase, developers instruct the system on how to distinguish and classify the different examples contained in the dataset. Third, during the deployment of the system, the statistical knowledge gained during the training phase, i.e., the model, is used to classify new inputs that were not initially present in the database, e.g., screening new text uploaded on a social media platform and classifying it as potentially glorifying terrorism.[8]

When used in anti-terrorism measures, machine learning can produce a disproportionate impact on marginalized and racialized communities. To understand these risks, identifying the biases present in machine learning systems throughout their entire life cycle is essential.

To begin with, biases may manifest in the questions the machine learning system is programmed to answer, how its deployment context is framed, or, how it is instructed to identify and classify the input received. Anti-Islamic social stereotyping that pervades the discussion around anti-terrorism significantly impacts such decisions, which, in turn, can lead the system to attach negative attributes to Islamic content to a greater extent than to, for example, Western White content.[9]

Biases also arise whenever the data forming the dataset is not fully representative of the populations or phenomena being modeled,[10] or if it mirrors historical and systemic biases,[11] such as those linking Islamic individuals and content to terrorism.[12] For instance, a system designed to screen social media group

[7] Hugo Verhelst et al., *Machine Learning Against Terrorism: How Big Data Collection and Analysis Influences the Privacy-Security Dilemma*, 22 SCI. ENG. ETHICS 2975, 2976.
[8] Stuart Russell & Peter Norvig, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH, 694-695 (3rd ed, 2010). Machine learning is a subset of artificial intelligence and consists of three sub-areas: supervised learning, unsupervised learning, and reinforcement learning. The described approach refers to supervised learning.
[9] Reva Schwarts et al., *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*, 16 (NIST Special Publication 1270, 2022) [explaining that the mere choice of these attributes (e.g., "criminality" or "trustworthiness") can, in light of the human intervention and its potentially biased identification, negatively impact the accuracy of the system]; Caroline Mala Corbin, *Terrorists Are Always Muslim but Never White: At the Intersection of Critical Race Theory and Propaganda,* 86:2 FORDHAM L. REV. 455 (2017) [explaining the unconscious cognitive biases that contribute to the creation of stereotypes and how such stereotypes are further incentivized by conscious narratives of white privilege]; Rachel R. Steele et al., *Bias Within Because of Threat From Outside: The Effects of an External Call for Terrorism on Anti-Muslim Attitudes in the United States*, 6:2 SOC. PSYCHOL. PERSONAL. SCI. 193 (2015) [explaining how Islamic extremist threats coming from outside of the United States increase anti-Muslim bias inside the US towards Muslim Americans].
[10] Simone Fabbrizzi et al., *A Survey on Bias in Visual Datasets,* arXiv preprint: 2107.07919 (2021), available at: https://arxiv.org/abs/2107.07919; Ninareh Mehrabi et al., *A Survey on Bias and Fairness in Machine Learning,* 54:6 ACM COMPUTING SURVEYS 1 (2021).
[11] Schwarts, *supra* note 9; ARTICLE 19, *supra* note 6 at 16.
[12] Abubakar Abid et al., *Persistent Anti-Muslim Bias in Large Language Models*, arXiv:2101.05783 (2021), available at: https://arxiv.org/abs/2101.05783 ; Sigal Samuel, *AI's Islamophobia problem*, VOX, (Sep. 18, 2021 8:00 AM),

memberships to predict the likelihood of an individual having extremist beliefs will disproportionately target Arabic-speaking individuals if Arabic data are disproportionally represented in the initial dataset.

Datasets used in training for counter-terrorism purposes are exceptionally challenging due to the nature of terrorist acts. Especially at the domestic level, terrorist acts are highly diverse in terms of individuals involved (i.e., whether it is an organization or a single individual), motives (e.g., whether the act is racially, ethnically, or religiously motivated), planning (i.e., whether it was an impulsive act or a highly-calibrated plan), and execution (e.g., whether it was a single isolated event or a series of simultaneous attacks).[13] This data diversity and scarcity (because of the disproportionate ratio of non-terrorist population and terrorist population[14]) can significantly lower the accuracy of machine learning systems.[15]

These issues are also present during the design, training, and validation phases.[16] In the field of counterterrorism, for instance, algorithms need to be trained to account for all the different kinds of information available for each example of the terrorist act in the dataset and find correlations among them. Due to the diversity among terrorist acts, the system is more likely to mismatch correlations in newer inputs. Lastly, stereotyping patterns are likely to be replicated when machine learning systems are designed to aggregate data about a group (e.g., an Islamic terrorist organization) to make inferred predictions about an individual (e.g., whether an individual who has Islamic beliefs is likely to represent a terrorist threat).[17]

4. Facial Recognition

Facial recognition softwares may be defined as technologies able to screen an image or video and recognize, identify, and categorize the individuals portrayed in it and, in some instances, the emotions such individuals are feeling through an analysis of facial expressions, including vocal tone, gait, and psychological signals.[18]

The advancements in facial recognition technology have made its use in national security, anti-terrorism, and crime prevention progressively more frequent over the past twenty years. National security agencies in the U.S. and in the European Union have disclosed that they are using facial recognition technology in border control and management and in national identification systems.[19] Human rights advocates have criticized such initiatives and pressured governments to impose a moratorium on the development and use of facial recognition technology until sufficient protections for individuals' freedoms are introduced.[20]

https://www.vox.com/future-perfect/22672414/ai-artificial-intelligence-gpt-3-bias-muslim ; https://hai.stanford.edu/news/rooting-out-anti-muslim-bias-popular-language-model-gpt-3

[13] Lasse Lindekilde et al., *Radicalization patterns and modes of attack planning and preparation among lone-actor terrorists: an explanatory analysis*, 11:2 BEHAV. SCI. TERROR. POLITICAL AGGRESS. 113 (2017); Matenia P. Sirseloudi, *How to predict the unpredictable: On the early detection of terrorist campaigns*, 21:4 DEF. SECUR. ANAL. 369 (2006).

[14] So-called *class imbalance* challenge. *See e.g.,* Alexandra L'Heureux et al., *Machine Learning With Big Data: Challenges and Approaches*, 5 IEEE ACCESS 7776 (2017).

[15] Verhelst, *supra* note 7 at 2978.

[16] *See e.g.*, Schwarts, *supra* note 9.

[17] Julia Angwin et al, Machine Bias: There's software used across the country to predict future criminals. And it's biased against black, ProPublica, (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; *see also* ARTICLE 19, *EU: Civil society urges EU to fix algorithms* (Sept. 22, 2021), https://www.article19.org/resources/eu-civil-society-urges-eu-to-fix-algorithms/.

[18] ARTICLE 19, *supra* note 6 at 8; Suja Palaniswamy et al., *Real-time emotion recognition from facial images using Raspberry Pi II,* (3rd SPIN, 2016).

[19] ARTICLE 19, *supra* note 6 at 9.

[20] ARTICLE 19, *supra* note 6 at 4.

These concerns are supported by claims from software developers themselves who have not yet validated the accuracy of facial recognition technology, as well as by increased media coverage of cases portraying its discriminatory outcomes towards racialized and marginalized individuals.[21] As for general machine learning applications, facial recognition technology has been proven to be far less accurate when analyzing non-white subjects and women, especially when such information is then used to derive the emotional status of the subject.[22]

Even if technological advancements were to allow facial recognition – and other machine-learning-based applications – systems' accuracy to be validated, their use would still be negatively impacted by the institutional bias inherent in the organizations deploying it.[23] Within national security, for example, guidelines for initiating and conducting investigations have been routinely loosened for actions concerning Muslim individuals, allowing broader surveillance and permitting racial profiling.[24]

5. Natural Language Processing

Machine learning can also be used to develop natural language processing systems aimed at digesting human spoken language to screen audio or textual information.[25] To be developed, natural language processing needs to be trained against datasets instructing it to associate certain words and phrases to, for instance, extremist speech or terrorism.

Datasets' systemic and historical biases – coupled with linguistic disparities – hinder the ability of natural language processing tools to correctly assess content by marginalized and racialized individuals.[26] The over-representation of English training sets and relative lack of training data in less digitally prominent languages poses serious limitations on natural language processing systems, generating higher error rates.[27] The diversity of linguistic variations in terms of accent, terminology, and ethnicity present in the United States creates further potential for misunderstanding and discrimination against non-English content.[28]

6. Algorithmic Screening Systems

Online, machine-learning-based screening systems are used to analyze and classify content that individuals upload. To do so, algorithmic screening brings together a set of tools that allow different kinds of media

[21] *See e.g.*, Alex Najibi, *Racial Discrimination in Face Recognition Technology*, SITN, (Oct. 24, 2020),https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/ ; Drew Harwell, *Federal study confirms racial bias of many facial-recognition systems, casts doubts on their expanding use*, WASH. POST, (Dec. 19, 2019 6:43 PM), https://www.washingtonpost.com/technology/2019/12/19/federal-study-confirms-racial-bias-many-facial-recognition-systems-casts-doubt-their-expanding-use/

[22] ARTICLE 19, *supra* note 6 at 9, 21-22.

[23] *See infra* points 6 and 7.

[24] *See generally*, Failza Patel, *Ending the 'National Security' Excuse for Racial and Religious Profiling*, Brennan Centre for Justice (Jul. 22, 2021) [discussing how the post-9/11 "permissive rules for intel-li-gence collec-tion, coupled with weak protec-tions for speech and against discrim-in-a-tion, have subver-ted legit-im-ate coun-terter-ror-ism aims"];

[25] See e.g., Christopher Manning and Hinrich Schutze, *Foundations of Statistical Natural Language Processing* (1999).

[26] Thomas Davidson et al., *Racial Bias in Hate Speech and Abusive Language Detection Datasets*, arXiv:1905.12516 (2019), https://arxiv.org/abs/1905.12516.

[27] *See e.g.*, Marteen Sap et al., *The Riskue of Racial Bias in Hate Speech Detection*, 1668 (Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019*)*.

[28] Natasha Duarte et al., *Mixed Messages? The Limits of Automated Social Media Content Analysis* (Conference on Fairness, Accountability, and Transparency, 2018), https://cdt.org/wp-content/uploads/2017/12/FAT-conference-draft-2018.pdf.

(text, photo, video, live-streamed content, etc.) to be classified. Examples of the tools involved include keyword filters and hashing technology (similar to digital fingerprinting).[29] The purpose of screening is to prevent the distribution of illegal and harmful content before it is made available online.

Algorithmic screening systems often lack the ability to understand context, which can lead to challenges distinguishing, for example, content which reports on or criticizes a terrorist act from material glorifying or promoting terrorism).[30] When algorithmic screening systems are developed, they are trained to apply statistical knowledge to assess and classify the input. Contextual clues, such as whether the author was a news reporter or an individual affiliated with a terrorist organization, are extremely difficult to translate into quantitative terms and, if unaccounted for, increase the likelihood of misclassifications.

7. Issues with surveillance technology used by governments

In addition to considering the technical challenges related to accuracy and discrimination, the use of surveillance technology by both governments and private technology companies needs to be weighed against its impacts on privacy.[31] Identifying the right balance is an extremely complex task.[32] Human rights experts have frequently stressed how inadequate or non-existent legal frameworks addressing the use of facial recognition, and data collection and processing technology, have enabled an arbitrary and disproportionate use of surveillance technology worldwide.[33] Such pervasive surveillance poses severe risks to individuals' expressive rights, both online and offline.

Mass surveillance has a well-documented harmful impact on the rights to freedom of expression and freedom of association. Behavioral studies have shown that individuals act differently when they are being observed[34] and self-police for fear of negative repercussions, both in the physical world and in their online behavior.[35] Individuals limit how they express themselves online, the groups and people they are associated

[29] Hashing technology is similar to digital "fingerprinting". Each piece of content uploaded online has its own "hash" which allows it to be quickly re-identify if re-uploaded. Hashing technology is mainly used in CSEA measures. *See generally*, Enrique Guerra & Bryce G. Westlake, *Detecting child sexual abuse images: Traits of child sexual exploitation hosting and displaying websites*, 122 CHILD ABUSE NEGL. 1.

[30] Joseph Cox & Jason Koebler, *Why Won't Twitter Treat White Supremacy Like ISIS? Because It Would Mean Banning Some Republican Politicians Too.*, VICE, (Apr. 25, 2019 6:21 PM), https://www.vice.com/en/article/a3xgq5/why-wont-twitter-treat-white-supremacy-like-isis-because-it-would-mean-banning-some-republican-politicians-too; Evan Enstrom & Nick Freamster, *The limits of fileting: a look at functionality and shortcomings of content detection tools*, (Engine, March 2017).

[31] Michael Cayford & Wolter Pieters, *The effectiveness of surveillance technology: What intelligence officials are saying*, 34:22 INFO. SOC. 88 (2018); Stephanie J. Bird, *Security and Privacy: Why Privacy Matters*, 19 SCI. ENG. ETHICS 669 (2013).

[32] See e.g., LAWFARE, *Snowden Revelations*, (2014), https://www.lawfareblog.com/snowden-revelations.

[33] OHCHR, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, A/HRC/23/40, (2013).

[34] Elton Mayo, THE HUMAN PROBLEMS OF AN INDUSTRIAL CIVILIZATION (1933) [demonstrating the Hawthorn effect, according to which workers' production increased parallel to the increase of attention focused on them]; Tim Eckmanns et al., *Compliance with antiseptic hand rub use in sensitive care units: the Hawthorne effect*, 27 Infect Control Hosp. Epidemiol. 931 (2006) [demonstrating that medical staff were 55% more likely to wash their hands when they were being watched]; Mark Warner & Victoria Wang, *Self-censorship in social networking sites (SNSs)-privacy concerns, privacy awareness, perceived vulnerability, and information management*, 17 JICES 375 (2009) [finding a positive correlation between the increase of self-censorship and the increase of privacy concerns].

[35] *See e.g.*, Jonathon W. Penney, *Internet surveillance, regulation, and chilling effects online: a comparative case study*, 6:2. INTERNET POLICY REV. 1.

with, what websites they visit, and even which products they shop for.[36] Facial and emotion recognition technology used to infer one's psychological state or to predict whether an individual could be a terrorist risk pose a range of threats to rights.[37]

Pervasive surveillance can also create an obvious chilling effect against protests and activism, particularly among individuals who have some fear of retribution.[38] When individuals are aware that they are being watched and tracked, their freedom of speech and association may be hindered by the fear of negative consequences.[39] Anyone merely participating in a protest can be identified through the combination of security images, social media, location data, messaging information, and sim card tracking.[40] When platforms use data collection and processing to profile individuals based on their online behavior, virtually all online expression becomes traceable, as the companies are able to infer volumes of private and sensitive information, such as one's religious beliefs, sexual orientation, and political ideology.[41]

Government and private companies' efforts to combat domestic terrorism that rely on predictive and identification technology can pose greater chilling effects on the expressive rights of marginalized, racialized, and minority groups. Law enforcement and national security agencies have a history of bias against Black and Brown subjects and other minorities, who are investigated and surveilled more often than White subjects. This results in a higher quantity of Black and Brown subjects' data in security databases, which are then used to train algorithmic systems, resulting in further over-representation and law enforcement focus.

This disproportionate targeting can lead to other harmful impacts. Journalism and activism are harmed since the awareness of being monitored and easily identifiable discourages their sources of information.[42]

[36] See e.g., Elizabeth Stoycheff, *Under Surveillance: Examining Facebook's Spiral of Silence Effects in the Wake of NSA Internet Monitoring*, 9 JMCQ 2 (2016); Jonathon W. Penney, *Chilling Effects: Online Surveillance and Wikipedia Use*, 31 BTLJ 1 (2016); Rafi Goldberg, *Lack of Trust in Internet Privacy and Security May Deter Economic and Other Online Activities*, (NTIA series on the results of the July 2015 CPS Computer and Internet Use Supplement, 2016).

[37] ARTICLE 19, *Emotional Entanglement: China's emotion recognition market and its implications for human rights,* 15-19 (Jan. 2021), https://www.article19.org/wp-content/uploads/2021/01/ER-Tech-China-Report.pdf

[38] Peter Ullrich & Gina Rosa Wollinger, *A surveillance studies perspective on protest policing: the case of video surveillance of demonstrations in Germany*, 3 INTERFACE 12 (2018) [discussing protestors' reactions to videotaping of political demonstration].

[39] London Policing Ethics Panel, *Final Report on Live Facial Recognition*, (May 2019), http://www.policingethicspanel.london/uploads/4/4/0/7/44076193/live_facial_recognition_final_report_may_2019.pdf ; EU FRA, *Facial recognition technology: fundamental rights considerations in the context of law enforcemen*t 20 (Vienna, 2020) 20.

[40] Albert Fox Cahn & Zachary Silver, *The Long, Ugly History of How Police Have Tracked Protestors*, FAST CO. (June 2, 2020), https://www.fastcompany.com/90511912/the-long-ugly-history-of-how-police-have-tracked-protesters; https://edri.org/our-work/we-need-to-talk-about-clearview-ai/ [on the issues with social media scraping used to build facial recognition softwares that are then used by law enforcement and national security agencies];

[41] Claudia Diaz, *On the Impact of Social Network Profiling on Anonymity,* in PETS: PRIVACY ENHANCING TECHNOLOGIES 44 (Nikita Borisov & Ian Goldberg eds., 2008).

[42] ARTICLE 19, *Press Freedom Under Threat*, 13-18, 22-24, (May 2018), https://www.article19.org/wp-content/uploads/2018/05/Press-Freedom-Under-Threat-International-Press-Freedom-Mission-to-the-United-States.pdf; Jay Mazoomdaar, *Delhi Police film protests, run its images through face recognition software to screen crowd*, The Indian Express, (Dec. 28, 2019); ARTICLE 19, *India: Tech firms should uphold privacy, free speech*, (Mar. 11, 2021), https://www.article19.org/resources/india-tech-firms-should-uphold-privacy-free-speech/

Content moderation systems are not exempt from the inherent biases discussed above, from upload screening filters to natural language processing, facial recognition, and human review.[43] Too often, systems are unable to understand the context and the linguistic nuances that might impact how certain words and phrases are perceived. This is the case, for instance, for specific terms which can be understood as hateful when used against marginalized groups but neutral when used among members of such a group.[44]

In anti-terrorism measures, companies' definitions of "terrorism", "act glorifying terrorist ideologies", or "violent extremism" are vaguely defined. Identifying whether the content is being uploaded to praise terrorist acts is an extremely complex and contextual task that generates relatively high error rates among screening systems.[45] Companies also rely on national and international terrorism identification lists to direct their designations of terrorist groups, which, in turn, guide content and account removals.[46] Content moderation systems match content uploaded and user profiles to the names and information contained in such lists to limit and remove the presence of individuals involved in terrorist activities. Studies analyzing companies' approaches to terrorist content have demonstrated an over-restriction of Muslim speech.[47] Over-removal and over-profiling do not only impact the expressive rights of the target individuals. Overly broad and opaque practices routinely result in the removal of journalist content due to screening systems' wrongful labeling,[48] undermining efforts to report and denounce illegal and harmful activities.

A disproportionate focus on Islamic extremist content also draws resources away from other significant threats.[49] For instance, content reflecting White supremacist ideologies has enjoyed comparatively weaker moderation.[50] Policies on white supremacy and white nationalism are more narrowly defined.[51] Rather than enforcing removals or other limitations on reach, platforms have instead opted for intermediate approaches such as labeling and downranking.[52]

[43]Davidson, *supra* note 24.

[44] Ángel Díaz & Laura Hecht-Felella, *Double Standards in Social Media Content Moderation*, Brennan Center For Justice 3, 11-12 (Aug. 4, 2021).

[45] Courtney C. Radsch, *On Christchurch Call Anniversary, a Step Closer to Eradicating Terrorism Online?*, JUST SEC. (May 21, 2021), https://www.justsecurity.org/76607/on-christchurch-call-anniversary-a-step-closer-to-eradicating-terrorism-online/.

[46] Faiza Patel & Mary Pat Dwyer, *So What Does Facebook Take Down? The Secret List of 'Dangerous' Individuals and Organizations*, JUST SEC (Nov. 1, 2021). *See also.,* Sanjey Sharma & Jasbinder Nijjar, *The racialized surveillant assemblage: Islam and the fear of terrorism,* 16 IJMC 1 (2018); Chinmay Arun, *Facebook's Faces*, 135 HARV. L. REV. (forthcoming); EU Commission, *Code of Conduct on Countering Illegal Hate Speech*, (May 20216), available at: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en ; Amar Tamoor, *France Wants Facebook and Twitter to Launch an 'Offensive' Against ISIS Propaganda*, THE VERGE (Dec. 3, 2015 11:38 AM); *ISIS Online: Countering Terrorist Radicalization And Recruitment On The Internet And Social Media*: *Hearing Before the Permanent Subcommittee on Investigation of the Committee on Homeland Security and Governmental Affairs,* 114th Cong (Jul 6, 2016).

[47] Cox & Koebler, *supra* note 28.

[48] Courtney C. Radsch, *GIFCT: Possibly the Most Important Acronym You've Never Heard Of*, JUST SEC., (Sept. 30, 2020), https://www.justsecurity.org/72603/gifct-possibly-the-most-important-acronym-youve-never-heard-of/ .

[49] ARTICLE 19, *Human Rights NGOs in Coalition Letter to GIFCT,* (Aug. 3, 2020), https://www.article19.org/resources/human-rights-ngos-in-coalition-letter-to-gifct/; Díaz & Hect-Felella, *supra* note 42 at 3.

[50] *Id.*

[51] Díaz & Hect-Felella, *supra* note 42 at 6.

[52] Cox & Koebler, *supra* note 28; Díaz & Hect-Felella, *supra* note 42 at 14-15.

8. Government involvement in content moderation

The risks posed to individuals' privacy, freedom of expression, and association are further exacerbated through indirect government involvement in private technology platforms' content moderation practices and policies. In the field of anti-terrorism, governments have gained access to private user data and demanded or prompted the removal of illegal material (including terrorist content) through mechanisms such as Internet Referral Units, codes of conduct, and other informal avenues of coordination with law enforcement.[53] Informal and voluntary initiatives encourage companies to err towards over-removal and put undue extra-legal pressure on social media companies to remove content.[54] This restricts freedom of expression and press freedom as it can negatively impact news media and journalistic content, particularly related to commenting on or reporting about terrorism.[55]

Government pressure to be faster and more effective at removing terrorist-related content can lead companies to take actions that harm user speech. The Global Internet Forum to Counter Terrorism (GIFCT) is a prime example. Within the GIFCT, member platforms coordinate content removal to limit the spread of terrorist content online.[56] Although each company can technically make its own content moderation decisions, many smaller companies defer to the decisions of larger ones. The lack of independent review and oversight means there is a lack of safeguards to ensure that these measures are appropriately targeted and tailored. The vast majority of content dealt with under GIFCT's moderation processes falls into the category of "glorification of terrorist acts", which is particularly ambiguous and likely to include collateral content, such as counterspeech or news reporting, and which ends up disproportionately impacting Muslim and Arabic voices.[57] Participating companies have justified the absence of a database of affected content with the liability that could arise in light of the limitations imposed by the EU General Data Protection Law. The United States' decision to join the GIFCT and the associated Christchurch Call makes it imperative to reconcile these challenges through the introduction of independent auditing, oversight, and review of their work in a manner which also respects privacy rights.

9. Other issues with public-private partnerships

Through public-private partnerships like the GIFCT, private technology companies are turned into a new type of enforcers of government policy, blurring the lines of public and private authority.[58] The delegation

[53] For instance, the Australian Sharing of Abhorrent Material Law, the EU Terrorism Content Regulation, and the German Network Enforcement Act, demand quick removal and reporting to avoid incurring financial and criminal liabilities. See, respectively, AUSTRALIA, Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019 No. 38, 2019; EU, Regulation 2021/784 Of The European Parliament and of the Council on addressing the dissemination of terrorist content online, OJ L172/79, 2021; Netzwerkdurchsetzungsgesctz [NetzDG] [Network Enforcement Act], July 12, 2017.

[54] *Cf,* Heidy Tworek & Paddy Laarsen, *An Analysis of Germany's Network Enforcement Act,* (TWG, Apr. 15, 2019).

[55] Courtney C. Radsch, *In Fight against Extremism, Press Freedom Must Not Be Compromised*, COMMITTEE TO PROTECT JOURNALISTS, (Feb. 20, 2015), https://cpj.org/2015/02/fine-line-between-countering-extremism-and-allowin/; Courtney C. Radsch, *Countering Violent Extremism and Media Development; An Uneasy Relationship, a Need for Dialogue*, Center for International Media Assistance, (Oct. 2016).

[56] Radsch, supra note 46.

[57] A human rights impact assessment of the GIFCT has found that its database embeds anti-Islamic bias, which reflects the concerns posed by human rights organizations denouncing the disproportionate focus on Islamist-linked extremist or terrorist content. See, BSR, *Human Rights Assessment: Global Internet Forum to Counter Terrorism*, (2021): https://gifct.org/wp-content/uploads/2021/07/BSR_GIFCT_HRIA.pdf ; ARTICLE 19, *supra* note 46.

[58] ARTICLE 19, *Side-stepping rights: Regulating speech by contract* (Jun. 19, 2018), https://www.article19.org/resources/side-stepping-rights-regulating-speech-by-contract/; Courtney C. Radsch,

of frontline content enforcement to private entities who act in coordination with, and often at the behest of, governments creates novel constitutional challenges and is detrimental to traditional models of speech regulation.[59]

When technology companies agree to share information with national security agencies, and vice-versa, through secretive agreements, individuals' privacy rights are undermined due to the lack of effective public oversight.[60] At present, there is no information available on the acquisition, trade, and use of private companies' technology and information by government agencies. This is further exacerbated when the sharing of information about individuals is not limited to potential threats to national security but, rather, extends to the entirety of the users of private companies' services, through, for example, the use of facial recognition systems that allow government agencies access to extremely invasive technologies without sufficiently strong procedural safeguards.[61]

Too often, there is no information on what technologies are being used and how, what government agencies are involved in their deployment, for which purposes, and whether risks posed to individuals' fundamental rights are sufficiently mitigated.[62]

10. Precedential Effect of U.S. Policy

The PCLOB should also recognize the role that the United States plays in shaping technology policy and its impact around the world. Foreign legal systems draw inspiration from the U.S. for domestic laws, as has happened in the past with the Patriot Act[63] and cybercrime legislation.[64] Also, many of the largest global private technology companies are based in the United States which means that their policies are primarily shaped by American laws, norms, policies, and social and political pressure.[65] In particular, major technology companies overwhelmingly reflect U.S. perspectives on expressive rights and freedoms in their global community standards, and are significantly more likely to change their behavior if pressured by U.S. institutions.[66] It is therefore imperative for the PCLOB to consider the extraterritorial impacts of U.S. efforts

Privatizing censorship in fight against extremism is risk to press freedom, CPJ (Oct. 16, 2015 10:17 AM), https://cpj.org/2015/10/privatizing-censorship-in-fight-against-extremism/ .

[59] Michael Karanicolas, *Subverting Democracy to Save Democracy: Canada's Extra-Constitutional Approaches to Battling "Fake News",* 17 CJLT 201, 215-22 (2019).

[60] *See e.g.,* Committee on Standards in Public Life, *Artificial Intelligence and Public Standards*, 52-56 (Feb. 2020) [noting that accountability requires public bodies' awareness of the risks posed by the AI systems they deploy]; ARTICLE 19, *supra* note 6 at 19-20.

[61] Sonja Solomun et al., *Submission to the House of Commons Standing Committee on Access to Information, Privacy and Ethics: Study on the Use and Impact of Facial Recognition Technology,* CENTRE FOR MEDIA, TECHNOLOGY AND DEMOCRACY (Jun. 2021), <https://perma.cc/3QWY-9Z8A> [discussing the issues stemming from government use of facial recognition technology].

[62] ARTICLE 19, *supra* note 6 at 19.

[63] Beth Elise, *Exporting the Patriot Act? Democracy and the 'War on Terror' in the Third World, 28* THIRD WORLD QUARTERLY 1017, 1023-1025 (2007).

[64] Markus Fallenböck, *The Digital Millennium Copyright Act*, the European Community Copyright Directive and Their anticircumvention Provisions, 7 IJCLP (2003), https://ciaotest.cc.columbia.edu/olj/ijclp/ijclp_7/ijclp_7d.pdf

[65] Blayne Haggart, *American Internet, American Platforms, American Values,* CIGI (May 5, 2021), https://www.cigionline.org/articles/american-internet-american-platforms-american-values/; Courtney C. Radsch, *On the Frontlines of the Information Wars- How Algorithmic Gatekeepers and National Security Impact Journalism*, In NATIONAL SECURITY AND JOURNALISM,(Marc Ambinder et al., eds, forthcoming).

[66] Natasha Tusikov, CHOKEPOINTS: GLOBAL PRIVATE REGULATION ON THE INTERNET (Univ of California Press 2016) [explaining that American platforms are more responsive to American legislators]. *See e.g.*, Elizabeth Thompson, *Ethics committee votes to subpoena Facebook's Mark Zuckerberg to testify*, CBC (May 7, 2019 6:35 PM),

to fight domestic terrorism, particularly when it comes to law and technology policy, given that jurisdictions with weaker rule of law safeguards risk being disproportionately impacted by American policies reflected in how private companies handle their global operations.

11. Conclusion

Countering the spread of extremist ideologies and preventing domestic terrorist acts is of utmost importance. Technology relying on machine learning has the potential to provide a substantial boost to efforts aimed at predicting and preventing terrorist acts. Likewise, the expansion of private sector influence over our online discourse necessitates robust engagement processes with major tech companies as part of these efforts. However, the growing role of the private sector and the increasing reliance on machine learning technologies both also pose significant risks, as outlined in this comment. These risks require the development of robust oversight and mitigation measures, and careful consideration to ensure that State actions are proportional. We invite the Board to take into consideration the observations made throughout this comment in crafting an appropriate balance between security and the protection of fundamental rights.

Respectfully Submitted,

Alessia Zornetta
Research Assistant
UCLA Institute for Technology, Law & Policy

Courtney Radsch
Resident Fellow
UCLA Institute for Technology, Law & Policy
U.S. Representative
ARTICLE 19

Vidushi Marda
Senior Programme Officer
ARTICLE 19

Michael Karanicolas
Executive Director
UCLA Institute for Technology, Law & Policy

https://www.cbc.ca/news/politics/facebook-zuckerberg-cambridge-analytica-1.5127007 [where Facebook's CEO refused to attend the International Grand Committee but have attended most US Congress Hearings when summoned].