The Promise Institute for Human Rights
UCLA School of Law

UCLA Center for Critical Internet Inquiry

HUMAN RIGHTS RACIAL EQUALITY & NEW INFORMATION TECHNOLOGIES

MAPPING THE STRUCTURAL THREATS

JUNE 2020

CONTENTS

The **Promise Institute for Human Rights at UCLA School of Law** and the **UCLA Center for Critical Internet Inquiry** convened an expert working group of leading international scholars of race and technology to discuss “**Human Rights, Racial Equality and Emerging Digital Technologies: Mapping the Structural Threats**.”

The in-depth discussions will inform a report by the **United Nations Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance, Tendayi Achiume,** UCLA Law Professor and Promise Institute Faculty Director, to be delivered to the **UN Human Rights Council** on **July 16, 2020**.

This report summarizes important highlights from a number of leading researchers who are experts in the specifics that inform this report, and who have written extensively at the intersection of race and technology.

With the rise of networked digital technologies, scholars and global internet rights advocates across different fields have increasingly documented the ways in which these technologies reproduce, facilitate or exacerbate structural racial inequality. However, global human rights discussion and mobilization against the harms of new information technologies mostly revolve around a specific set of issues: hate crimes and hate speech, the role of new information technologies in facilitating hate incidents online, the use of online networks and fora to coordinate racial hatred, and threats to individuals’ right to privacy. This work is undoubtedly vital, and ensuring that governments address racist and xenophobic speech, incitement to hatred and acts of violence serves an important purpose in attaining racial equality. But just as important is a broader structural analysis and response to the intersection between emerging networked and predictive technologies, from software and hardware to internet platforms, and their effect on racial equality.

With this in mind, the workshop was structured around three themes:

(i) mapping the political economy and other structural forces that drive patterns of racial discrimination and exclusion around the world;

(ii) developing a typology of the most pressing forms and mechanisms of structural racial discrimination and inequality associated with emerging digital technologies, including algorithmic oppression, discrimination and bias, and automated and interactive machine learning for public and private decision-making affecting communities and individuals through a variety of institutions and practices;

(iii) outlining the appropriate human rights legal, policy and advocacy responses rooted in global human rights, civil rights, sovereign rights, equality and non-discrimination norms, including reform and accountability proposals, and highlighting areas in need of conceptual or theoretical development.

The following highlights some of the central themes that arose during the daylong discussions.

MAPPING THE POLITICAL ECONOMY

Jessie Daniels, Chris Gilliard, Jasmine McNealy and Hayley Ramsay-Jones offered opening remarks to frame the discussion.

Our entire society is now technological, which has profound impact and implication for racial equality. Technology is skewed by the biases of those that create it: yet there is tremendous denial on the part of Silicon Valley elites and governments to acknowledge and act upon the myriad social harms that emanate from their products and services, most of which have no oversight by the public vis-à-vis regulatory and legislative agencies and policymakers. Further, as Jessie Daniels explained, the language of digital technology is American English and networked technologies are imbued with American norms, giving the United States tech industry disproportionate economic and political power, and outsized influence among global multinational corporations of all types, often reflecting imperialist and white supremacist ideologies (Daniels, 2018), which include deep investments in narratives of technology companies’ commitments to colorblindness.

Hayley Ramsay-Jones emphasized that the biases of tech creators are embedded at every level of the tech design process, from conception to production and distribution. Consequently, there is no way to have an unbiased algorithm, as they are fundamentally flawed by design (Noble, 2013, 2014, 2018; O’Neil, 2016). The application of these algorithms leads to the amplification of “otherness” and to the neglect and exclusion of groups that do not look like the creators (powell & Menendian, 2016). Depending on the application, racial biases can result in relatively benign outcomes such as advertising Afro-Caribbean hair products to Afro-Caribbean women, to far more lethal applications such as fully autonomous weapons targeting people of color. Sarah Roberts highlighted that tech is constructed to focus on and seek out different types of activities and behaviors that have been pathologized and encoded as abhorrent by those with power. Furthermore, claims that algorithms and artificial intelligence can be “unbiased” or optimized for ethics in the new “ethical AI” movement has been challenged by critical information studies scholars and critical race and digital studies scholars as more of the same techno-determinism that has created the current conditions of crisis (Noble, 2018).

Tech infuses every facet of our lives, from education and banking to employment, housing and health care, and more (O’Neil, 2016; Noble, 2018; Benjamin, 2019). In each of these sectors, personal data is collected, individuals are monitored, and everything is informationalized (Schiller, 2007). Chris Gilliard explained that tech has changed human interaction by utilizing digital platforms to displace human-to-human contact in decision-making. Standardization on such a broad scale is problematic because it has the effect of monetizing exchanges, atomizing individuals and disintegrating community, leading to dehumanization and exclusion. Elana Zeide emphasized that lack of standardization across various software systems used for social decision-making (e.g., criminal sentencing software and other predictive analytics) make it difficult to track and audit harm across various implementations. Jasmine McNealy explained that biased algorithms lead to biased outcomes, amplifying the harm and difficulties that marginalized groups experience, while reducing decisional diversity and the ability for human intervention to take individual circumstances into account, because most of these systems are opaque and the public or lawmakers or system users are unable to assess the potential for negative impact until after the harms are done.

In addition, Gilliard suggests that technologies of these types are increasingly being deployed by private citizens and companies and, when collated with government-collected data, enables a combined web of information, which would not be possible if only one of these groups had access. This interrelationship between the state and the industry is indicative of the collapse of the nation state as tech companies rise in global power, where internet platform policy becomes a form of soft power on the international stage.

Despite the incredible potential for tech to harm individuals, there are very few laws placing limits on the ways in which tech is used by companies, institutions or states to monitor, gather data and make decisions. The policies of social media and other tech ecosystems and products are designed to protect companies' interests, rather than to protect users. The entire industry is driven by profit, allowing for gross market manipulation enabled by computational tools. Furthermore, tech is easily manipulated by the influence of dark money, which encourages the spread of dis- and mis-information (Nadler et al., 2018).

While the early logic of the pre-commercial internet was that it would be hyper-democratic and foster broad and equal participation and access, the reality 50 years later is quite different. We have more people online globally than ever due to large-scale commercial internet platforms, even though there continues to be a digital divide between the Global North and South. In this context, American-founded internet giants dominate much of the global technology landscape; exporting with them a variety of American-infused ideas about privacy and free speech that are defined in the interests of the executives and managers of these companies, which do not represent racial, ethnic or gender diversity. Instead, a small technocracy made of primarily white American men set the governance and ideological dispositions that lead, inform, guide, and translate a set of values that are increasingly a threat to civil and human rights in the United States, and beyond. In this context, big tech hastens the speed and scale with which discourses and structures are spread, allowing discrimination to be multiplied and amplified at a pace never seen before, with relatively little international governance or policy oversight.

DEVELOPING A TYPOLOGY

Simone Browne, Chaz Arnett, Margie Cheesman, Dragana Kaurin and Bryan Mercer offered opening remarks to frame the discussion.

Two particular groups are at significant risk of harm from racial discrimination exacerbated by the use of new technologies: those involved in the criminal justice system, and refugees and asylum seekers. Structural inequalities also stem from labor extraction of those involved in the production and disposal of tech (Simone Browne) and from algorithmic help-seeking in the context of sexual discrimination (Kate Sim).

Electronic projects of containment rebrand and repackage incarceration in a variety of forms. Machine-based decision-making is used for risk assessment in the ongoing struggle to eliminate cash bail, using algorithms that draw on data points from criminal data sets, as well as data sets gathered from outside the criminal justice system, to make predictions on both where crime may happen in the future, and whether an individual will fail to appear for a bail hearing in the present. Bryan Mercer emphasized that the data inputted into machine learning algorithms used for such predictive analytics is low quality, shaped by and reflective of biased policing practices which disproportionately target people of color, resulting in deep racial disparities with high false-positive rates of risk for people of color (French & Browne, 2014).

Alternative means of incarceration, for example through the use of ankle monitors, subjects individuals to continuous surveillance. E-carceration is branded as something new when in fact it is a continuation of a biased system that pushes a false narrative of safety when the real agenda is targeted control of individuals (Arnett, 2019; Browne, 2015). Like mass incarceration, e-carceration is disproportionately used against people of color, with its impact being felt along race and class lines, pushing those under surveillance further into the

margins of society (Arnett, 2018; Benjamin, 2019). As with the data used in bail predictions, careful research shows how data sets are biased, and in addition, machine modeling and algorithms create proxies for race by using other factors, such as number of arrests. This perpetuates the racial hierarchy in the carceral state, without checks and balances available against the algorithm and its implementation and management.

Refugees and asylum seekers are increasingly required to provide personal and biometric data for identification, access to services, and other purposes. These data are collected by a variety of actors, including law enforcement, border patrols, and NGOs. Consent to collection is assumed and there is a lack of alternatives to providing the data (Kaurin, 2019). Margie Cheesman argued that biometric proliferation is a form of function creep whereby consent is assumed rather than given when refugees are expected to interact with biometric interfaces for essential services beyond identity registration. She argued that informed consent should include meaningful choice and therefore involve alternative options, e.g. ID cards instead of biometrics. Migration industry actors should consider whether new technologies address the most pressing problems in refugees' lives, and the inequalities they extend (Sim & Cheesman, 2020). Kaurin suggested data agency as an alternative framework to consent, emphasizing that, in many cases, refugees' autonomy over their digital bodies is extremely limited. This is because paternalistic logics underpin inadequate efforts to address issues of data rights and risks with refugees. The severe risks include data leaks, monitoring and persecution. Kaurin stressed that these individuals are often unable to make informed decisions because they don't know where their data will be stored, who it will be shared with and what the possible threats are, further eroding their individual agency and potentially their human rights, as data collected may have longer term consequences in relationship to the distribution of food, shelter, medical attention, and other crucial social goods (Kaurin, 2019). Kaurin and Cheesman both questioned the validity of current consent frameworks for data exchanges, highlighting the uneven power structures refugees find themselves in.

They argued that addressing digital inequalities require greater transparency, channels of accountability, and the representation of refugees in design and decision-making around digitalization projects and data exchanges (Kaurin, 2019; Sim & Cheesman, 2020).

Kaurin and Cheesman both argued that refugee identification technologies are not neutral tools to provide rights (such as access to food, medical care, etc.), but are part infrastructures of mobility control and surveillance (Kaurin, 2019). Likewise, mobile technologies and social media platforms present both an opportunity and a threat for refugees as they are used as a crucial means of information and communication, but are also tools of profiling and surveillance by governments and private actors (Gillespie, Osseiran & Cheesman, 2018).

Kaurin explained that individuals in the system often withhold data that would be beneficial to their asylum claim because they do not understand or trust how the data will be used, and fear that information they provide might get back to those they are fleeing.

OUTLINING A HUMAN RIGHTS RESPONSE

Tamás Kádár, Matthew Bui, Jessica González, Meetali Jain and Hamid Khan offered opening remarks to frame the discussion.

Among the experts on race and technology convened to inform this report, there is broad agreement that technology is both infused with a variety of troubling values—intentionally or not, and these technologies are also deployed in a variety of global contexts that render them uneven and unstable, particularly as they are used by and against vulnerable communities (Arnett, 2018, 2019; Browne, 2015; Eubanks, 2018; Noble, 2018; Benjamin, 2019; Broussard, 2019; Daniels, 2019). Technology is a social practice imbued with social relations, and this is where critical information studies scholars, for example, foreground both the implications of technology design, but also the ways in which it is governed and managed in a variety of social, economic, and political contexts (Browne, 2016; Buolamwini & Gebru, 2018; Eubanks, 2018; Noble, 2018). Particularly in the United States, much discourse around technology at the regulatory level is predicated on false framings of what the purpose of a variety of these technologies are and who the customer is. These have been purposely miscast by the companies creating the technology in the pursuit of profit, with the effect of obscuring the discriminatory impact of technology.

In order to function, algorithms have to ascribe some type of “universal person” or “ideal” as the model, encoding a binary classification between “right” (the “ideal” person) and “wrong.” These binary classifications perpetuate hierarchies of power, as the “wrong” equates to “otherness,” i.e. those without the power and privilege of those designing and training the algorithm.

Put another way, all models are predicated upon an ideal that makes up the model or optimal person or condition (which may be constituted by thousands of data points), and deviations from that ideal model are typically disadvantageous in the prediction or outcome the model is classifying. What research is showing us is where deviations from the model or ideal in a large-scale predictive system has a host of racist, sexist, and class-based consequences. In addition, algorithmic systems create more nuances for systematically disadvantaged groups by incorporating proxies for race and gender, which can be difficult for humans to recognize in large scale big data computing contexts (Browne, 2012; Noble, 2018). Indeed, many narrow-AI tasks use thousands of data points that may be deeply inaccurate, but by the nature of large-scale big data computing, they by definition, are looking for patterns and to make predictions that are imperceptible to humans (O'Neil, 2016). This raises a number of ethical and moral considerations about the ability for humans to audit algorithms and AI that are reliant upon big data computational frameworks. Indeed, the complete lack of transparency in the creation and implementation of new forms of technologies allows for the continued development of discriminatory algorithms, as there is no way to challenge the biases of the creators and the inadequacy of the biased data used to train the algorithms (Burrell, 2016). Finally, as Tamás Kádár underscored, there are very few transparent processes that exist by which algorithmic decision-making can be challenged by those affected, adding another layer of opacity to their impact and to legal and civil remedy.

CONCLUSION

Combating these challenges will require a multi-faceted approach, encompassing the technology platforms themselves, governments, and investigations to enhance transparency. Governments should be called upon to increase regulation of tech companies, and government institutions involved in algorithmic development, to strengthen civil rights and privacy rules to protect individuals. The international and interconnected nature of technology raises significant challenges to ensure that individuals' information is protected by tech companies and on servers outside of their country of nationality, and will require significant inter-governmental coordination and cooperation. Furthermore, individual nation-states should ensure that tech companies are taxed to reflect the role that they play in the market.

In addition to governmental intervention, Jessica González called for the bringing together of groups from human rights, civil rights, sovereign rights, and digital rights to collaborate to pressure big tech companies to change their policies to become more protective of users. In particular, consumers of tech products have immense power to push for change and to force the framing, rather than waiting for big tech to take the initiative, and Hamid Khan highlighted the proven power of community action at the grassroots level in pushing for change.

In making these changes, there are questions about where the tensions exist between a host of technologies, some of which are more reliant upon machine learning and algorithms (or narrow-AI), and some of which rely heavily upon practices like commercial content moderation to address speech and the proliferation of illegal or harmful content. Many large-scale digital platforms are contending with all of these issues. The questions of the responsibility for tech company practices and where due diligence should lie are questions legislative bodies are asking of large internet media companies and should be supported and even coordinated through the United Nations.

Imperative questions must be asked, such as: Should we trust tech companies to police themselves or should there be external monitoring bodies? What could external monitoring bodies look like, and who would take part?

Finally, deeper philosophical questions exist about the overall desirability of algorithmic decision-making that completely removes human involvement from decisions that impact people's lives, which we will continue to study and convene to explore.

APPENDIX A - CONVENING PARTICIPANTS

Organizers

- Tendayi Achiume, UN Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance; Professor of Law, UCLA School of Law; Faculty Director, Promise Institute for Human Rights
- Kate Mackintosh, Executive Director, Promise Institute for Human Rights, UCLA School of Law
- Safiya Umoja Noble, Associate Professor, UCLA Department of Information Studies and African American Studies; Co-Director, UCLA Center for Critical Internet Inquiry
- Sarah T. Roberts, Associate Professor, UCLA Department of Information Studies
- Kai Fees, Research Affiliate, Promise Institute for Human Rights; Advisor to the UN Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance

Presenters, Participants & Observers

- Chaz Arnett, Assistant Professor of Law, University of Pittsburgh School of Law
- Simone Browne, Associate Professor, University of Texas at Austin
- Matthew Bui, Doctoral Candidate, USC Annenberg School for Communication
- Margie Cheesman, Digital Anthropologist, Oxford Internet Institute
- Jessie Daniels, Professor, Hunter College and The Graduate Center, the City University of New York

- Fanna Gamal, Binder Clinical Teaching Fellow, UCLA School of Law
- Chris Gilliard, Professor of English, Macomb Community College
- Jessica González, Co-CEO, Free Press
- Meetali Jain, Legal Director, Avaaz
- Tamás Kádár, Deputy Director (Head of Legal and Policy), Equinet
- Dragana Kaurin, Human Rights Researcher; Founder, Localization Lab
- Hamid Khan, Campaign Coordinator, Stop LAPD Spying Coalition
- Ruth Livier, Doctoral Candidate, UCLA Department of Information Studies
- Jasmine McNealy, Associate Professor, University of Florida; Fellow, Stanford University Digital Civil Society Lab
- Bryan Mercer, Executive Director, Media Mobilizing Project
- Jessica Peake, Director, International and Comparative Law Program; Assistant Director, Promise Institute for Human Rights, UCLA School of Law
- Hayley Ramsay-Jones, Director, Soka Gakkai International Office for UN Affairs
- Kate Sim, PhD Researcher, Oxford Internet Institute
- Rebecca Tsosie, Visiting Professor, UCLA School of Law
- Elana Zeide, PULSE Fellow in Artificial Intelligence, Law and Policy, UCLA School of Law

APPENDIX B - PARTICIPANT OBSERVATIONS

JESSIE DANIELS

The need for critical race theory for understanding the racial harms of the algorithmic internet

When John Perry Barlow declared that he and others on the early Internet would “create a civilization of the Mind in Cyberspace,” where “our identities have no bodies,” in 1996 there were already white supremacists colonizing online space. The contemporary adherents of Barlow’s race-less vision of technology continue to hold sway in domains from tech policy to computer science to social theory about the Internet. This supposedly race-less perspective is a form of colorblind racism (Daniels, 2015). The colorblind racism of the tech industry, and of theorizing the web, represents one of the greatest impediments to addressing the spread of racism and discrimination online.

Background: Internet Studies & Critical Race Theory

Since Barlow’s manifesto and the earliest iteration of what was then called the world wide web, and sometimes “the information super highway” or the “electronic frontier,” of the early 1990s, there have been two subsequent iterations of the popular Internet, Web 2.0 (1999-2007) and the current algorithmic Internet (2007-2019). Taking up the algorithmic Internet of social media, machine learning and artificial intelligence and the way it reinforces structures of racial inequality even as it fuels extremist racist violence. Yet, the hegemonic understanding of the Internet as “race-less” is a form of epistemological ignorance, as Charles Mills conceptualizes it, that makes it impossible to understand the world that Silicon Valley has created.

Social Media Platforms & the Algorithmic Spread of Racism

Social media platforms both ignore racism in their design and thereby leave the platforms open to exploitation by white supremacists. I refer to these white supremacists as “innovation opportunists.” (Daniels, 2018). Following each innovation in media and/or technology, white supremacists look for openings to exploit in order to spread their ideology. The people who run the platforms and, often the researchers and journalists who write about them, continue to express surprise at the “bug” of white supremacist content, while evidence suggests that it is a “feature.”

Algorithms and Systemic, Structural Racial Inequality

There are three key research monographs in the last several years that detail the way algorithmic Internet reproduces systemic, structural racial inequality. In Weapons of Math Destruction (2016), Cathy O’Neil charts the development of algorithms as “just math,” without considering race, yet that nevertheless reproduce racial inequality in financial services, housing and education. In Algorithms of Oppression (2018), Safiya Noble examines the way that the algorithms used by search engines reproduce the racist terms entered by users, thus amplifying and speeding up harmful ideologies as search results. And, in Race after Technology (2019), Ruha Benjamin argues that algorithms hide, speed up, and deepen racial discrimination through technologies that make race part of the architecture of everyday life.

Each of the laws meant to address racial discrimination in the United States Civil Rights Acts of the 1960s, in voting, housing, hiring/employment, and public accommodations, has been disrupted by digital media technologies, and made much worse by algorithms.

Responses: Abolitionist Technologies & Advancing Racial Literacy in Tech

Ruha Benjamin (2019) proposes “abolitionist” technologies for addressing what she calls the New Jim Code. But the path to abolitionist technologies is not clear when the tech industry itself is rife with racial microaggressions, facilitates the spread of white supremacist ideologies and attendant violence, and seems impervious to attempts at even the most minimal diversity in hiring, promotion and power sharing.

In one intervention, described as a “harm reduction approach to the problem of microaggressions," is racial literacy. In a 2019 report, Daniels, Nkonde and Mir describe this approach as having three components: 1) cognitive – educating oneself about race and racism; 2) emotional – learning to deal with racially stressful situations; and 3) a commitment to take action to create an organization committed to anti-racism. Without a racial literacy approach, United States-based tech companies will remain stuck in the cul-de-sac of unconscious bias trainings that let those who perpetrate racial microaggressions off the hook.

JESSICA GONZÁLEZ

There is widespread consternation about the threat online platforms pose to civil rights, racial justice, and safety. Online platforms have hastened the spread of hate and misinformation. They have exploited people’s personal data and private information in myriad ways with little accountability, transparency or consequence. They’ve worsened the crisis in journalism while spreading misinformation.

Officials, advocates and editorial pages everywhere are clamoring to do something — but there’s little consensus on what exactly that something should be. The levers for change seem inadequate or obscure. Many existing policy proposals are either weak tea or dangerous cures worse than the disease they are supposed to treat. And all the while our eyeballs stay glued to our devices as these companies keep getting richer, growing bigger, and becoming more integrated into people's everyday lives.

Three core threats that online platforms pose are: (1) the spread of misinformation paired with a rapid decline in quality journalism lead to a less informed electorate and make it hard to hold power accountable; (2) the amplification and normalization of hate and bigotry are leading to real life violence and the normalization of cruel and inhumane policies and beliefs on immigration, criminal justice, etc.; and (3) privacy abuses and online discrimination threaten people's civil and human rights.

Free Press is working on several key strategies to address these threats.

Antidote To Misinformation: Platform Tax To Support The News We Need

As online platforms have grown, reliable and independent journalism that people need to engage in constructive dialogue and sort out their differences continues to disappear from communities. Yet as audiences increasingly move toward reading news filtered through online platforms, traditional advertising is no longer a sustainable source of funding for journalism.

This is the problem: Too little online advertising money is flowing to content with high social value.

Quality investigative journalism and independent reporting are rife with what economists call "positive externalities," meaning the benefit to society is greater than the benefit to those who directly access or pay for the content. Conversely, online hate, trolling, misinformation and hyper-commercialism create "negative externalities" by harming society in ways that do not always directly affect the content producer, consumer or platforms that distribute this content.

Free Press believes the best approach to addressing the platform's contributions to the crisis in journalism may be an old one — a tax. In this case, a tax would be levied on the targeted-advertising business model itself and redistributed to fund the high-value, democracy enhancing journalism that's missing. Think of it like a carbon tax imposed on the oil industry to pay for cleaning up the mess it has made. We should do the same to the targeted-advertising industry to counteract how the platforms amplify content that is polluting our civic discourse.

Lawmakers should tax the purveyors of targeted-advertising and put the revenues in a public trust fund to support production and distribution of content by diverse speakers — with an emphasis on local projects, investigative reporting, media literacy and journalism produced by and serving people of color and other underserved communities. There are many details to be worked out (and there is more than one way to design such an effort). But a tax of 2 percent on targeted-advertising could produce more than $2 billion per year.

Antidote to Online Hate: Broad Coalition Pressure To "Change The Terms"

Internet platforms' core algorithms are designed to gather people into like-minded groups and feed them the content that creates the strongest reaction. White-supremacist organizations worldwide have used online platforms to organize, fund, recruit, and normalize and promote racism, sexism, xenophobia, religious bigotry, homophobia and transphobia, and to coordinate violence and other hateful activities. These attacks chill the online speech of the targeted groups, frustrating democratic participation in the digital marketplace of ideas and — even more importantly — threatening their safety and freedom in real life. We aim to help disrupt these coordinated efforts over social media, web hosting services, and financial-transaction websites.

Not every pressing online platform problem has a legislative or policy fix. While the First Amendment limits the United States government's role in policing speech, it does not apply to these private platforms. People have no inherent right to algorithmic amplification or to promote racism, xenophobia and other forms of bigotry on online platforms. We believe companies like Facebook, Google and Twitter have a responsibility to protect users and confront how their platforms are used to spread hate.

Together with Center for American Progress, Southern Poverty Law Center, and dozens of human and digital rights groups, Free Press launched Change the Terms in October 2018, a coalition which aims to pressure the platforms to curb and reduce hateful activities online.

The group developed model corporate policies that balance the values of stopping hateful activities online while upholding the values of free expression, due process and transparency. The goal is to get online platforms and financial-transaction companies to adopt corporate policies that prevent the spread of hateful activities and follow procedures to ensure those policies are enforced in a transparent, equitable, culturally relevant way by a team that comes from impacted communities, with clear and easy ways to appeal any decisions. The model policies cover terms of service, enforcement, transparency, evaluation, governance and more.

So far Change the Terms has influenced changes at platforms large and small, and the coalition is growing in visibility and power.

Antidote To Civil Rights And Privacy Violations In The Digital Age: Legislation

Tech companies have used our data to enable and sometimes even participate in discrimination against people of color, women, members of the LGBTQ community, religious minorities, people with disabilities, immigrants and other marginalized communities.

We must ensure that powerful interests don't use our data in ways that violate our rights and silence our voices. We must have control over how our personal information is used, and prohibit its use to build systems that oppress, discriminate, disenfranchise and exacerbate segregation.

Together with Lawyers' Committee for Civil Rights Under Law, Free Press has drafted a model legislation to:

- Outlaw algorithms that discriminate in employment, housing, retail and lending, voting and public accommodations;
- Give people far more transparency about how their data is being used, along with a private right of action to sue companies that violate their digital civil rights;
- Prevent apps from surreptitiously tracking us across the web;
- And more.

TAMÁS KÁDÁR

The most pressing forms and mechanisms of structural racial discrimination and inequality associated with new information technologies

New information technologies (new IT) have a significant potential to exacerbate or lead to racial discrimination and inequalities. These impacts might be involuntary or they can be consciously sought for.

It is important to note that machines do not discriminate of their own accord. The people creating and using new IT hold the same biases and stereotypes that lead to the 'traditional' acts of discrimination in society and new IT, particularly AI, reproduces discrimination. Stereotypes regarding some ethnic groups viewed by police as more prone to committing crimes affects the way preventive policing is developed and used, disproportionately targeting Roma people and persons of African descent, for example.

In order to effectively and systematically analyze the different forms and mechanisms of racial discrimination and inequalities stemming from the use of new IT, one could usefully categorize them according to the fields of life affected. Given the wide, and quickly extending, use of new IT, examples of discriminatory effects can be found in a plethora of fields. This includes:

- Employment (e.g. the use of new IT in recruitment or career progression or in displaying job ads to different audiences);
- Education (e.g. in the allocation of university places);
- Goods and services (e.g. as regards financial and insurance products; discriminatory price differentiation of online shops; targeted online advertising);
- Housing (e.g. in the allocation of public housing);
- Healthcare (e.g. concerning health insurance or in decisions concerning medical treatment);
- Social protection (e.g. in the allocation or monitoring of social benefits);
- Policing (e.g. in predictive policing);
- Immigration control (e.g.at border control);
- Justice (e.g. assessing the likelihood of recidivism);
- Hate speech in public (online) space.

Three more important aspects merit attention.

First, new IT by its very nature often has a transnational component (e.g. in online hate speech cases or when a particular software is developed by a foreign company) that makes it unusual compared to some of the well known and documented examples of "traditional" discrimination.

Second, a lack of transparency (also dubbed as "black box effect") is characteristic to new IT and it has the potential to render discrimination invisible both to the rights-holders and the institutions empowered to remedy it.

Third, it is important to acknowledge that while new IT carries significant risks, it also holds an important potential for tackling discrimination and inequalities.

It is important that these potential advantages of new IT are publicly discussed, noted and used to the best possible extent. Such positive potential of new IT can include for example:

- Using new channels to raise awareness about situations of inequality, the risks of new IT, as well as equal treatment laws and institutions (such as equality bodies) that can provide effective remedies;
- Collecting evidence in discrimination cases using the ability of new IT to analyse and digest large quantities of (e.g. statistical) information that would otherwise be impossible or extremely human resource-intensive;
- Conducting research to identify and expose inequalities and assisting scientific innovations supporting the achievement of substantive equality for vulnerable individuals (e.g. reasonable accommodations at the workplace and in educational settings for persons with disabilities).

Appropriate responses to the challenges posed by new technologies

First, it is important to be vigilant and to react to the dangers to inequality and equal treatment caused by new IT, but we need to be mindful that in certain respects, the problems caused by new IT are also subject to some hyping and could be presented in a disproportionate way. This is particularly the case with regard to the appropriate responses to inequalities and discrimination caused by new IT. A lot of the issues in question can be tackled with the (legal and institutional) means already available. In many ways, therefore, the main question appears to be how to make sure that the existing framework is adjusted and implemented in order to allow it to effectively apply to the new problems.

In this process it is important to acknowledge and react to specificities of new IT such as the reliance on an immense amount of data or the role of multinational corporations. This points to the importance of re-thinking and revising traditional ways of dealing with discrimination issues, looking beyond classical anti-discrimination law and ensuring a good use of other legal avenues such as data protection rules (when new IT is handling large amounts of data in breach of data protection rules), competition law (when large multinational corporations might be abusing their dominant positions), or consumer protection law (when consumers are subject to manipulative on-line advertising). It is also important that the burgeoning field of digital rights advocates and stakeholders recognize the importance of equality (and its transversal, cross-sectoral nature) as a key concern in their work and are introduced to the variety of complementary legal tools for tackling AI-driven discrimination.

If we are to provide appropriate responses to the challenges posed by new IT to equality and non discrimination, it is paramount to ensure that sanctions in cases of discrimination are effective and dissuasive. While this is a horizontal issue for anti-discrimination law in general, the large number of potential victims of discrimination makes it even more imperative to ensure that sanctions reflect the societal risks of discrimination by and with new IT. Individual cases of discrimination brought by one person, where the decision only compensates the damages suffered by that individual will not be sufficient to deter users of new IT and to remedy the damages caused to society. Therefore, collective and class actions and punitive damages could prove even more necessary for new IT discrimination cases than for other, "traditional" types of cases. The introduction of legal mechanisms such as mandatory public sector equality duties (legal duties on the public sector to have due regard to equality in carrying out their functions) would ensure that government as an increasingly prominent user of AI technologies is held accountable for its impact on equality in society.

In new IT, the strong international components and strong corporate actors wielding massive influence mentioned above will necessitate cooperation among states and between states and corporate actors. Initially, impetus for such cooperation could be given by effective and genuine self-regulation, monitored by responsible state actors. Ultimately, however, self-regulatory principles and a focus on ethics and AI

should not overshadow the importance of strong and effective legislation and enforcement: ethics is no substitute for human rights.

Last, but not least, in the face of new and large-scale challenges posed by new IT, it is important to ensure the strengthening of the institutional framework for equality and against discrimination. The lack of oversight in the development and use of new IT will exacerbate discriminatory effects. Equality bodies, as public institutions specialized to tackle inequalities and discrimination, are important actors in this field (alongside others) and are already dealing with different forms of discrimination and inequalities caused by new IT. However, the scale and the novelty of the challenges posed by new IT require that they are accorded the competences (including powers to collect evidence and to litigate), expertise (either in-house or external) and the resources (both human and financial) to be able to effectively respond to discrimination by new technologies.

HAMID KHAN

For too long the analysis of state violence and its impact has been comfortably rooted in the soft narratives of constitutional violations, few bad apples, need for more training, more diversity etc. These arguments continue to guide and control the debate and advocacy shared amongst the “progressive” advocacy community and its cohorts in the media and non-profit world. Such arguments not only miss core issues but dangerously continue to create an illusion of rights hence leading us down the un-ending fight for “reform and accountability.” Furthermore, the invasion of privacy and violation of civil liberties narrative sorely misses and undermines clear analysis that the police state is an ever-expanding endeavor which is fundamentally and inherently flawed by design, intended and organized to repress and control Black, Brown and poor communities causing irreparable physical and emotional harm.

Our work has to offer deeply enriching and provocative understanding and analysis that expose multiple trajectories of the national security police state including the development, legitimization, and operationalization of tools of social control. While surveillance and data gathering were always an integral part of policing, the information revolution and the unholy marriage between policing and the post 9/11 reconfiguration of national security has led to an unprecedented expansion of both. Behavioral surveillance and data mining have become the primary modes of speculation and hunches under the guise of "pre-emptive" policing. Furthermore, such tools of social control are not limited to law enforcement but are deployed through many sectors such as social services, health care, housing and employment. It is the surveillance and policing of our bodies in every aspect of our lives. Communities of color, immigrants and the economically marginalized are the primary targets of the new modes of surveillance. These ever-expanding regimes of monitoring and control often unfold under the color of law. Consequently, critiques and resistance to these regimes remain imprisoned in legal frames of reference and reformist agendas. Our work has to offer an alternative framework for critique and resistance by exposing the expansion of police surveillance in the inherent structural imperative of violence and control – the foundational logic of law enforcement.

The applications and enforcement of such tactics are not limited to law enforcement but permeate all sectors of our society. Some of the tactics and programs include: Incorporation and codification of counter-terrorism and counter-insurgency tactics into domestic policing; Programs such as the National Suspicious Activity Reporting (SAR) originally intended for counter-terrorism intelligence is now a routine local policing practice under the guise of "all crimes" approach; Predictive Policing, which is grounded in counter-insurgency mapping on the battle fronts of Afghanistan and Iraq is rapidly becoming local policing methodology in "crime prevention" strategies; Predictive algorithms are being incorporated into social services to identify "abusive parents" in child protective services.

Such technologies are also being applied in private sector for "weeding out" problem tenants. Counter-terrorism practices like See Something, Say Something are being replicated by apps such as nextdoor.com by residents of upscale neighborhoods for identifying "suspicious" individuals; Electronic surveillance technology built for military use on the battle front such as facial recognition and bio-metrics collection or cell phone catchers, also known as stingray, are being increasingly incorporated for "investigative" purposes; vague and abstract concepts like "observed behavior" and "reasonable indication" are becoming key determinants in crime fighting, legitimizing hunch-based and speculative policing; Programs such as Countering Violent Extremism have led to the creation of FBI guidelines for all schools around the country to identify "problem" youth.

Our challenge is to advance current thinking by locating ever-expanding multi-sector surveillance of marginalized and vulnerable communities as a complement and facilitator of police violence and incarceration. Our practice must turn the focus of resistance struggles from legalistic police reform to abolition of policing as we know it and re-direct resources toward communities' self-sustenance. Our organizing requires us to intersect with communities of color, immigrants, economically marginalized, youth led organizations, liberation movements, cultural warriors, community organizers and opinion-makers.

In order to fully understand the impact of the national security police state it is incumbent upon us to meticulously map multiple audiences layered within and outside social justice movements. The range of this mapping should include those who actively seek community partnerships with law enforcement and conduct themselves as "shock absorbers" of the system to grassroots organizers and community members who fight for the abolition of our perpetual carceral conditions but remain marginalized and considered "rigid" or "fringe" even in progressive circles.

JASMINE MCNEALY

Amplifying “Otherness”

In their 2016 essay on inclusiveness and belonging, john a. powell and Stephen Menendian called “othering” “[t]he problem of the twenty-first century.” Defining othering as a system and structure that marginalizes and perpetuates inequality based on categories of identity, including religion, sex, and race, among other things, the scholars identified political and social conditions and power dynamics that promote group based othering in the world. The human tendency toward categorization and unconscious bias helped to explain the dynamics of othering; segregation, secession, and assimilation were failed responses to othering or the problem of the other. powell and Menendian proposed inclusion and belonginess – “unwavering commitment to not simply tolerating and respecting difference but to ensuring that all people are welcome and feel that they belong in the society” – as the way forward.

I agree with powell and Menendian’s assessment of othering, and argue that emerging information technology is amplifying otherness through neglect, exclusion, and disinformation, all of which have significant consequences. Neglect, while perhaps the most recognized problem with emerging technology, is persistent. By neglect I mean the creation, use, and deployment of technology without regard for the disparate impacts they may have on communities different than the imagined audience. Ignorance of the effects of technology can be both intentional and unintentional. Unintentionality presumes a developer did not know or had not considered the possible impacts of their technology. Creators embed their creations with their own values, and values reflect culture and politics. If communities are outside of the scope of the creator’s purview, they may fail to recognize the consequences of that technology for that community.

More insidious, perhaps, is intentional neglect, when in the creation, use, or deployment of technology the impact on a community is both known and ignored. A readily available example of this amplified othering through neglect is the implementation of algorithmic decision-making systems in the criminal justice process in the United States. Though touted as a way to circumvent bias in human decision-making in pretrial and sentencing, these machine learning systems are trained on data reflecting societal biases and systemic racism in the American criminal justice system. And although organizations creating these systems are aware of the biases in the training data, and the consequences, they continue to sell these systems to state and local governments, which then deploy them on their constituents. Whether neglect is intentional or unintentional, then, the discriminatory impact on communities of people should not be acceptable.

Exclusion, keeping particular groups from participating in various ways, is a significant impact of amplified othering. Algorithmic decision systems, like those mentioned above, are more likely to exclude members of some communities from full participation based on biased historical data. Not only are these systems deployed in the United States criminal justice system, but also in the financial sector, where they are used to decide whether an institution should extend credit for home or business loans. Such systems have also been shown to have discriminatory impact when deployed in human resources systems in choosing candidates to interview for jobs, as well as candidates for graduate and professional schools. Unlike the unconscious bias that powell and Menendian discuss in their essay, developers, scholars, journalists and others are now aware of the biases in these systems. Yet, adoption continues. This may be a reflection of persuasive framing communications used by governments and organizations creating and implementing these technologies despite public outcry.

Indeed, persuasion through framing is a part of language. Individuals and organizations persuade us to accept particular meanings and interpretations by making certain aspects of an idea more salient than others. Advances in communications technology have allowed the persuasive messages of disinformation campaigns to swell around the world, amplifying otherness, and resulting in race, gender, sexuality, and other identity-based violence. Social media manipulators are able to obscure the source of false information, while convincing those with a significant audience to propel their misleading messages. As a result, a larger audience may encounter deceptive communications, which may increase the vulnerability of certain communities. Social media disinformation campaigns have been identified as abetting the genocide of Rohinyans in Myanmar and influencing elections in Kenya, Brazil, and the United States, among other countries. Emerging disinformation technology is amplifying othering in additional ways. Deepfakes technology, for example, allows the user to make it appear as though an individual is saying or doing something they have not said or done. Because of the severe ramifications this technology on our political systems and for those targeted, legislators are considering passing laws.

But can technology-specific laws change othering? Certainly, legislation aimed at banning particular uses of technology and the deployment of harmful technology on the public is welcome. The recent successful campaigns to ban the use of facial recognition technology in San Francisco, Somerville, Massachusetts, and Oakland, for example, are important to helping to push back against government surveillance and the disparate impacts of those activities. But even more impactful would be the passage or strengthening of laws aimed at remedying othering and its historic and current impacts. Voting rights, gender equity, fair pay, and comprehensive privacy/data governance legislation, among other things, and the enforcement of these laws would go a long way in helping to remedy the underlying social issues amplified in emerging technology. While we may, and should, prohibit the use and deployment of harmful technology, it is important that we use law to manifest the belongingness and inclusion powell and Menendian identify – “that all people are welcome and feel that they belong in the society.”

BRYAN MERCER

Pretrial Risk Assessment Tools as a case of high-stakes machine-based decision making re-embedding structural racism against Blacks and other United States people of color

Reflections compiled from excerpts of a forthcoming website resource from Media Mobilizing Project and MediaJustice on Pretrial Risk Assessment Tools with contributions from Jenessa Irvine, Di Luong, and Hannah Sassaman of Media Mobilizing Project

In hundreds of communities across the United States, courts are embedding risk assessment algorithms into high-stakes pretrial incarceration, supervision, and release decisions. Their use is on the rise at the same time that social movements and communities struggle to end the use of money bail, and massively reduce pretrial incarceration. Most of the time the way these tools are used and the bias they embody is not obvious to the public. Media Mobilizing Project's research with hundreds of jurisdictions across the country works to make clear to communities both how and where these risk assessments are used and why they do not always lead to their stated goals of making our pretrial system smaller or fairer.

Pretrial Risk Assessment Tools (RATs) are computer algorithms that are supposed to help judges decide if someone should be released before their trial. They input variables about someone after they have been arrested and produce a risk score. Variables include elements from their criminal history, like previous convictions, as well as demographic factors such as age and housing stability. Pretrial RATs typically produce outcomes for Failure to Appear (FTA) and Recidivism, either as separate scores, or combined into one score. These scores translate into risk levels, which are then used to help judges assign release conditions or decide to detain someone pretrial. Most often, pretrial RATs produce a risk score for failure to appear in court (FTA) and/or for recidivism or new criminal activity (NCA).

RATs inform judges at an essential moment in the criminal process — whether or not someone gets to come home before their trial can have a huge impact on the ultimate outcome of their case. Studies have shown that being stuck in jail awaiting trial can result in longer sentences as well as a higher likelihood of conviction, often because people plead guilty to get out of jail. Being detained can also mean someone loses their job, homes, and children.

RATs as Algorithm Decision Making without Accountability

Some propose risk assessments as the necessary step in removing cash bail. We reject the idea that cash bail, a system that targets and harms poor people and people of color, needs to be replaced with another system that does the same thing.

There are many different RATs used across the country. Each state has its own laws governing pretrial assessments; some states have one standardized tool in use, while others allow counties and cities to choose which tool, if any, they will use.

There is a severe lack of clarity and transparency around what goes into these tools and how they are used — especially when tools are proprietary.

People charged with crimes have a right to know what is being used to evaluate them and how it impacts what happens to them. There is no lawyer present to represent a defendant at a risk assessment, and not necessarily any chance to review a score or contest it before a bail hearing.

Plus, those who use the results of algorithms may or may not understand how scores are calculated, or even know all the inputs that go into the tool to create the outcome. They will not necessarily know what particular inputs pushed an overall score to be especially high or low.

For example, if someone is arrested and they receive a high pretrial risk score, it may be inflated because of their age or because of their multiple previous convictions.

Even though these may indicate very different things about a person, all a judge may see is that they were determined to be “high risk”, without any context.

RATs give a false sense of scientific fairness, when in reality fairness is a human decision based on human bias.

Many of these algorithms have a well-documented disproportionate impact based on race; they draw from racially biased data and racially biased systems, so they reproduce inequalities in their outcomes. For instance, many tools use demographic factors such as whether or not someone is employed or has stable housing, which are both clearly connected to economic status.

No tool is created or used in a vacuum. The factors used, how they are weighed, and even what defines risk are all subject to human bias. Removing this decision to a computer merely bypasses the responsibility.

Validation scores and statistics show that the RATs used are not necessarily highly accurate or predictive. In many places, the tools are not validated at all or have not been tested on local populations. As with most oppressive systems, these impacts fall primarily on people of color and poor people. That is why we want to make sure that the pretrial risk assessments judging people, helping determine if they are free or not before their trial, do not uphold and reproduce the same biases already embedded in our criminal legal system.

HAYLEY RAMSAY-JONES

Racism and Fully Autonomous Weapons

The rise of artificial intelligence is largely due to an increase in power, memory and speed of computers, and the availability of large quantities of data about many aspects of our lives. Through the commercial application of big-data, we are increasingly being sorted into categories and stereotypes. In its most benign form this categorization is being used to sell us products via targeted advertising. However, in its

.

most egregious application we see the weaponization of new information technologies utilize similar categorizations based on biased algorithms, to which the consequences for certain communities could be deadly.

In this paper I focus on fully autonomous weapons that are currently being developed for military and law enforcement purposes, and their potential threat to the human rights of marginalized communities — in particular, persons of color interjectionally. This paper will also consider the systemic nature of racism and how racism would be reinforced and perpetuated by fully autonomous weapons.

Racism in Artificial Intelligence

Fully autonomous weapons can select and attack targets without meaningful human control; they operate based on algorithms and data analysis programming. In essence, this means that machines would have the power to make life-and-death decisions over human beings.

The trend towards more autonomy in weaponry without adequate human oversight is alarming especially when we know that digital technologies are not racially neutral. Moreover, when it comes to artificial intelligence (AI) there is an increasing body of evidence that shows that racism operates at every level of the design process and continues to emerge in the production, implementation, distribution and regulation. In this regard, AI not only embodies the values and beliefs of the society or individuals that produce them but acts to amplify these biases and the power disparities.

One example of racism manifesting in AI is the under representation problem in STEM fields, which in itself is a product of racism and patriarchy in western society, and the educational system. Technologies in the West are mostly developed by white males, and thus perform better for this group. A 2010 study by researchers at NIST and the University of Texas, found that algorithms designed and tested in East Asia are better at recognizing East Asians, while those designed in

Western countries are more accurate at detecting Caucasians. Similarly, sound detecting devices perform better at detecting male, Anglo-American voices and accents as opposed to female voices and non-Anglo American accents.

Research by Joy Buolamwini reveals that race, skin tone and gender matter when it comes to facial recognition. Buolamwini demonstrates that facial recognition software recognizes male faces far more accurately than female faces, especially when these faces are white. For darker skinned people, however, the error rates were over 19%, and unsurprisingly the systems performed especially badly when presented with the intersection between race and gender, evidenced by a 34.4% error margin when recognizing dark skinned women.

Despite the concerning error rates in these systems, commercially we already see adaptations of faulty facial recognition systems being rolled out in a variety of ways from soap dispensers to self-driving cars. The issue here is what happens if law enforcement and national security become reliant on a system that can recognize white males with just 1% error rate, yet fails to recognize dark skinned women more than one third of the time?

These types of applications of new information technology fail people of color intersectionally at a disturbing rate. The fact that these systems are commercially available reveals a blatant disregard for people of color; it also positions “whiteness” as the norm, the standard for objectivity and reason. These applications of new information technology including their weaponization favors whiteness at the expense of all others; it is not merely a disempowerment but an empowerment. In real terms, “racism bolsters white people’s life chances” (Reni Eddo-Lodge).

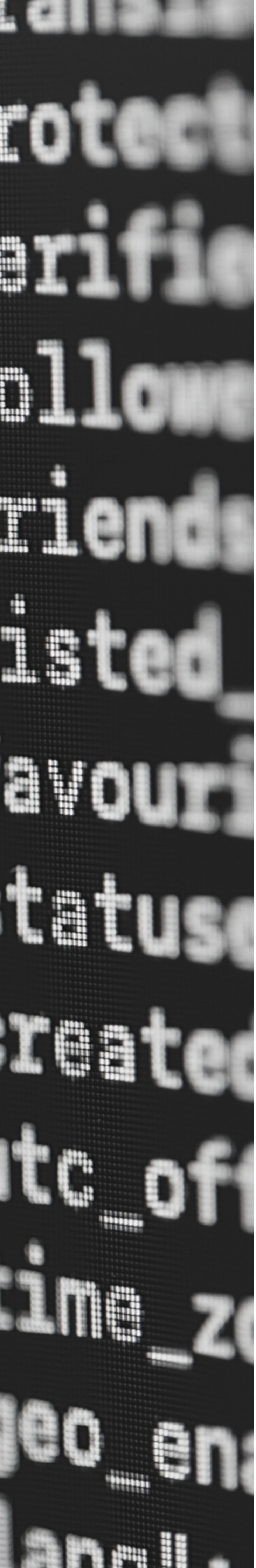
Historical or latent bias in data is another issue. This is created by frequency of occurrence, for example if you google image search “professional hair,” images of hair styles of mostly white women appear. Conversely, if you search “unprofessional hair,” images of mostly black women with afro-Caribbean hair emerges. This is due to machine learning — algorithms collect the most frequently submitted entries and therefore reflects statistically popular racists sentiments. These learnt biases are then projected into future search results, thus the racism continues to reproduce itself.

A more perilous example of this is in data-driven, predictive policing that uses crime statistics to identify “high crime” areas and then subjects these areas to higher and often more aggressive levels of policing. Crime happens everywhere, however when an area is over-policed, such as communities of color, then that results in more people of color being arrested and flagged as “persons of interest” — a self-fulfilling prophecy.

In 2017, Amnesty International launched a report called “trapped in the Matrix.” The report highlighted racially discriminatory practices by the United Kingdom police force and their use of a database called the “Gangs Matrix,” which inputs data on “suspected” gang members in London. As of October 2007, there were 3,806 people on the Matrix: 87% of those are from black, Asian and minority ethnic backgrounds, and of that percentage 78% are black — a disproportionate number given that the police’s own figures show that only 27% of those responsible for serious youth violence are black.

Amnesty stated that some police officers in the United Kingdom have been acting like they are in the “*Wild West*,” making false assumptions about people based on their race, gender, age and socio-economic status. As a result, individuals on the Matrix database are subject to chronic over-policing, with black people six times more likely to be stopped and searched than white people, and ten times more likely to be convicted of drug-related crimes.

This system not only interferes with their right to privacy. Amnesty claims that the police often share the Matrix with other local agencies such as job centers, housing associations, social services, schools and colleges. In several cases, this has led to devastating impacts on people’s social and economic lives because they are listed as “nominal” gang members, a label which is deliberately vague and stigmatizing.



The nature of systemic racism means that it is embedded in all areas of society; the effects of this type of oppression doesn't easily dissipate. Through the continual criminalization and stigmatization of people of color, systemic racism operates by creating winners and losers regardless of what people actually do. This is also the way that it redistributes opportunities and resources based on nothing other than privilege.

Given that the United Kingdom, as well as five other countries, are developing fully autonomous weapons to target, injure and kill based on data-inputs and pre-programmed algorithms, we can see how longstanding inherent biases pose an ethical and human rights threat. Where some groups of people will be vastly more vulnerable than others, fully autonomous weapons would not only act to further entrench already existing inequalities but could exacerbate them and lead to deadly consequences.

Legalities

As AI technology advances, the question of who will be held accountable for human rights abuses is becoming increasingly urgent. Machine learning and AI, affect a range of human rights including privacy, freedom of expression, freedom of assembly, the right to non-discrimination and equality, the right to life and the right to human dignity.

Holding those responsible for the unlawful killings of people of color by law enforcement and the military is already a huge challenge in many countries, however this issue would be further impaired if the unlawful killing was committed by a fully autonomous weapon. Who would be held responsible: the programmer, manufacturer, commanding officer, or the machine itself? As well as the above human rights concerns, the issue of proportionality and accountability in international humanitarian law should render these weapons unlawful as it would be difficult to obtain justice for the victims and their families.

Conclusion

According to Reni Eddo-Lodge, racism perpetuates partly through malice, carelessness and ignorance; it acts to quietly assist some, while hindering others. It is within this framework that we must grapple with race and the weaponization of new information technologies. In this regard, we should ask ourselves who controls these technologies and what do they think they know about the populations they are “categorizing”? What are the politics of these relationships and the deeply rooted systemic forms of discrimination? Who benefits from these technologies and how?

There is a long history of people of color being experimented on for the sake of scientific advances from which they do not benefit. An example of this is from James Marion Sims, known as the father of gynecology for reducing maternal death rates in the United States in the 19th century. He conducted his research by performing ghastly experiments on enslaved black women. “All of the early important reproductive health advances were devised by perfecting experiments on black women,” (Harriet A. Washington). Centuries later, the maternal death rate for black women in the United States is three times higher than it is for white women.

Thus, when it comes to new information technology — facial recognition systems, algorithms, and automated and interactive machine decision-making — communities of color are often both underserved by their benefits and overexposed to their consequences. This dual bind, where communities of color are subjected to science rather than supported by it, must be addressed.

We must guard against building deeply rooted social problems further into our technical infrastructure. We must work towards a zero policy on racism in tech, and not weaponize racism in tech. “We must take a precautionary approach to the use of AI in weapons technology and in fully autonomous weapons in particular” (Noel Sharkey). For these reasons, the Campaign to Stop Killer Robots, numerous governments, regional groups, tech workers, experts, scholars and the UN Secretary General are all calling for a prohibition treaty against fully autonomous weapons.

KATE SIM

This written reflection contributes to the UN Special Rapporteur's report on systemic inequalities and injustices by taking the design and implementation of data/AI-driven reporting systems for sexual harassment and violence in the United States and the United Kingdom. The contribution draws mostly from three-years of ethnographically informed research on sexual assault reporting apps in United States higher education to highlight the following: access and exclusion; bias in design; autonomy, confidentiality, and privacy.

> Following months of racialized and gendered harassment from her colleague, Hana wants to explore reporting options and support resources. Hana recalls her employer advertising an online workplace misconduct reporting system and decides to give it a try. Halfway through using the form, Hana realizes there is no way to document the kind of harassment she experienced, because the platform requires her to submit a single incident and choose only one category of harassment. Hana manages to complete the form, only to discover that it has been shared with her supervisor. Hana had only wanted to discuss her options at this stage and feels pressured to make a formal complaint. Disappointed and overwhelmed, Hana ultimately decides against seeking further help and her workplace experience continues to suffer.

The scenario above highlights how the digitization and automation of help-seeking procedures exacerbate issues of racial and gendered inequalities and injustices. Data- and AI-driven systems increasingly mediate the help-seeking, reporting, and evidence collecting experience of sexual harassment and violence. From algorithmic sexual assault reporting platform to AI chatbots for workplace harassment and intimate partner abuse evidence collection app, emerging technologies claim to provide privacy, objectivity, and neutrality. To institutions grappling with the ethics of internal misconduct resolution, these systems are a welcome intervention. To victims who have reasons to distrust their institutions, third-party systems are an appealing alternative.

However, as technology scholars studying digital inequalities and injustices have demonstrated (Broussard, 2019; Eubanks, 2018; Noble, 2018), such systems are far from accurate, objective, or neutral. In the context of sexual harassment and violence, victims' control over disclosure, how it is received, and what kind of information and resources they are directed to (Ahrens, 2006; Holland & Cortina, 2017) have a serious impact on their sensemaking of what happened and ability to trust others (Brison, 2003; Smith & Freyd, 2014; van der Kolk, 2015). In my doctoral research, I examine how the datafication and automation of help-seeking process raise the following ethical, social, and political implications. By "help-seeking," I refer broadly to the documentations, disclosures, evidence collection, and in/formal reports involved in responding to incidents of sexual harassment and violence. For purposes of this workshop, I propose algorithmic help-seeking as a companion concept to growing scholarship on algorithmic bias, discrimination, and decision-making (Kim, 2016; Kroll et al., 2016; Myers et al., 2019). The aim is to examine how computational rhetoric of objectivity (Broussard, 2018) and optimization (Burrell, 2016) instructs the design and application of data/AI-driven systems for sexual harassment and violence reporting.

Access and exclusion: Who is the intended user and how are they invited to use data/AI-driven systems to report sexual harassment and violence? Who is left behind? In the case of workplace misconduct reporting systems, the software assume the user to be a full-time employee of a physically shared workplace. This assumption leaves out temp, sub-contracted, and service workers even though they are more likely to experience discrimination and harassment (Yeung, 2019). As a result, these systems magnify the differential resources available to victims based on their race, economic class, and employment status, among others.

Bias in design: What are the biases about discrimination and harassment built into the design of these systems? What kind of data do these systems privilege and, by extension, what experience of victimhood do these systems privilege? Consider campus sexual assault reporting systems like Callisto and LiveSafe. The reporting interface's misconduct categories and

demographical information collected (or not collected) assume the user's whiteness, heterosexuality, and experience in a single incident of physical/sexual violence. Feminist scholars across law, history, sociology, and philosophy, (Freedman, 2013; Fischel, 2016; Grigoriadis, 2017; Yap, 2017) have long examined the racialized and gendered bias of assuming the innocence of white and male offenders and the culpability of victims of color. Uncovering the bias informing the design of data/AI-driven reporting systems highlights how such assumptions are value-laden in ways that map onto existing patterns of bias.

Evidence and the politics of credibility: Many of these systems are designed with the intention of affording the victim-user credibility through strategic uses of data. Some even provide instructions on how to gather digital and/or physical evidence, and automatically generate a testimonial. However, these systems and the advice they give are provided by system vendors with little or no legal expertise. In my interviews with practitioners and legal professionals, vendors' perception of credibility in the civil/criminal court is often deeply misguided and can seriously mislead users.

Autonomy, confidentiality, and privacy: Who has access to these systems? What levels of control are afforded to the user? Literature on mandatory reporting (Brodsky, 2018; Holland et al., 2018) strongly advises against forced disclosures and identify mandatory reporting policies as detrimental to victims' recovery. Some victim-users are likely to explore reporting systems with the understanding that their uses remain private and informal. When these disclosures are made accessible to the system vendors and/or authorities without their consent, it creates a serious breach of confidentiality and trust.

Considering the ethical, social, and political implications above highlights how the computational rhetoric of objectivity and optimization undergirds the design and adoption of reporting systems. As these systems increasingly mediate victims' help-seeking experience in ways that remove individual agency and silo victims based on their social group, it is all the more urgent to generate frameworks, policies, and practices to interrogate their applications.

APPENDIX C - SELECTED BIBLIOGRAPHY

- Ahrens, Courtney E. "Being Silenced: The Impact of Negative Social Reactions on the Disclosure of Rape." *American Journal of Community Psychology* vol. 38, no. 3–4 (December 2006): 31–34. doi.org/10.1007/s10464-006-9069-9.
- Amnesty International. *Trapped in the Matrix: Secrecy, stigma, and bias in the Met’s Gang Database*. 2018.
- Arnett, Chaz. *Race, Surveillance, and Resistance*. Ohio State Law Journal (Forthcoming 2021).
- Arnett, Chaz. *From Decarceration to E-carceration*. 41 Cardozo L. Rev. 641 (2019).
- Arnett, Chaz. *Virtual Shackles: Electronic Surveillance and the Adultification of Juvenile Courts*. 108 J. Crim L. & Criminology 399 (2018).
- Asaro, Peter. *Will #BlackLiveMatter to RoboCop?*. robots.law.miami.edu/2016/wp-content/uploads/2015/07/Asaro_Will-BlackLivesMatter-to-Robocop_Revised_DRAFT.pdf.
- Benjamin, Ruha. *Race After Technology: Abolitionist Tools for the New Jim Code*. New Jersey: John Wiley & Sons (2019).
- Bingham, Laura. *People v . Côte d’Ivoire: The Right to Citizenship for Minorities*. Retrieved from Open Society Justice Initiative website (2019): www.justiceinitiative.org/litigation/people-v-c-te-divoire.
- Brison, Susan. *Aftermath: Violence and the Remaking of a Self*. Princeton, N.J.: Princeton University Press (2003).
- Brodsky, Alexandra. *Against Taking Rape ‘Seriously’: The Case Against Mandatory Referral Laws for Campus Gender Violence*. Harvard Civil Rights 53 (2018): 131–66.
- Broussard, Meredith. *Artificial Unintelligence: How Computers Misunderstand the World*. MIT Press (2019).
- Browne, Simone. *Dark Matters: On the Surveillance of Blackness* (2015).
- Browne, Simone. "Digital Epidermalization: Race, Identity and Biometrics." *Critical Sociology* 36(1) (2010): 131-150.
- Browne, Simone. "Race and Surveillance." *Routledge Handbook of Surveillance Studies*. Eds. Kirstie Ball, Kevin D. Haggerty, and David Lyon. Routledge (2012): 72-79.

- Bui, Matthew, and Moran, Rachel. *Making the 21st century mobile journalist: Examining definitions and conceptualizations of mobility and mobile journalism within journalism education*. Digital Journalism. doi: 10.1080/21670811.2019.1664926, 2019.
- Bui, Matthew, and Noble, Safiya U. (forthcoming). "We're missing a moral framework of justice in artificial intelligence: On the limits, failings, and ethics of fairness," In M. Dubber, F. Pasquale, and S. Das (Eds.), *Oxford Handbook of Ethics of Artificial Intelligence*. Oxford: Oxford Univ Press.
- Buolamwini, Joy, and Gebru, Timnit. *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. Proceedings of Machine Learning Research 81:1-15, 2018.
- Burrell, Jenna. *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*. Big Data & Society 3, no. 1 (January 5, 2016): 205395171562251. doi.org/10.1177/2053951715622512.
- Daniels, Jessie, Nkonde, Mutale, and Darahkshan Mir. *Advancing Racial Literacy in Tech*. Report, New York, Data & Society (2019).
- Daniels, Jessie. "Racism in Modern Information and Communication Technologies." *Hunter Publications and Research* (2019), academicworks.cuny.edu/hc_pubs/495/.
- Daniels, Jessie. *The Algorithmic Rise of the "Alt-Right."* Contexts 17, no. 1 (2018): 60-65.
- Daniels, Jessie. "My Brain Database Doesn't See Skin Color" Color-Blind Racism in the Technology Industry and in Theorizing the Web." *American Behavioral Scientist* 59, no. 11 (2015): 1377-1393.
- Daniels, Jessie. "Twitter and White Supremacy: A Love Story." *Hunter Publications and Research* (2017). academicworks.cuny.edu/hc_pubs/344/.
- de Vreese, Claes H. "News Framing: Theory and Typology." *Information Design Journal & Document Design* 13, no. 1 (2005): 51-62.
- Donovan, Joan, and Friedberg, Brian. *Source Hacking: Media Manipulation in Practice*. New York, NY: Data & Society Research Institute (September 4, 2019). datasociety.net/output/source-hacking-media-manipulation-in-practice/.
- Eddo-Lodge, Reni. *Why I'm No Longer Talking to White People About Race.* (2018).

- Escobar, Arturo. *Encountering Development: the making and unmaking of the Third World*. Princeton: Princeton University Press (1995).
- Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press (2018).
- Fischel, Joseph. *Sex and Harm in the Age of Consent*. ProQuest Dissertations Publishing (2016). search.proquest.com/docview/894262381/?pq-origsite=primo.
- Freedman, Estelle . *Redefining Rape*. Harvard University Press (2013).
- French, M., and Browne, Simone, "Surveillance as Social Regulation: Profiles and Profiling Technology." *Criminalization, Representation, Regulation: Thinking Differently About Crime*. Eds. Deborah Brock, Amanda Glasbeek, and Carmela Murdocca. University of Toronto Press (2014): 251-284.
- Gardner, Katy, and Lewis, David. *Anthropology and Development; Challenges for the Twenty-First Century*. London: Pluto Press (2015).
- Gillespie, Marie, Osseiran, Souad, and Cheesman, Margie. *Syrian Refugees and the Digital Passage to Europe: Smartphone Infrastructures and Affordances*. Social Media and Society, (SI: Forced Migrants and Digital Connectivity Syrian) (2018). doi.org/10.1177/2056305118764440
- Grigoriadis, Vanessa. *Blurred Lines: Rethinking Sex, Power, and Consent on Campus*. Houghton Mifflin Harcourt (2017).
- GSMA. (2017). *Refugees and Identity: registration and aid delivery*. Retrieved from gsma.com/mobilefordevelopment/wp-content/uploads/2017/06/Refugees-and-Identity.pdf.
- Holland, Kathryn J., and Cortina, Lilia M. "'It Happens to Girls All the Time': Examining Sexual Assault Survivors' Reasons for Not Using Campus Supports." *American Journal of Community Psychology* 59, no. 1–2 (March 2017): 50–64. doi.org/10.1002/ajcp.12126.
- Holland, Kathryn J., Cortina, Lilia M, and Freyd, Jennifer J. "Compelled Disclosure of College Sexual Assault." *American Psychologist* 73, no. 3 (April 2018): 256–68. doi.org/10.1037/amp0000186.
- Kaurin, Dragana. *Data Protection and Digital Agency for Refugees*. (2019), www.cigionline.org/publications/data-protection-and-digital-agency-refugees

- McNealy, Jasmine, Nah, S., Kim, J.H., Joo, J. *Communicating Artificial Intelligence: Theory, Research and Practice.* Communications Studies (forthcoming 2020).
- McNealy, Jasmine, Cooper, N., Poblet Balcell, M. “Blockchain for Good: Special Issue on Democracy and Civic Technology.” *Frontiers in Blockchain* (forthcoming 2020).
- Moran, Rachel, and Bui, Matthew. *Race, ethnicity, and telecommunications policy issues of access and representation: Centering communities of color and their concerns.* Telecommunications Policy 43(5), 461-473 (2019). doi: 10.1016/j.telpol.2018.12.005.
- Nadler, Anthony, Crane, Matthew, and Donovan, Joan. *Weaponizing the Digital Influence Machine: The Political Perils of Online Ad Tech*. Data and Society (October, 2018). datasociety.net/output/weaponizing-the-digital-influence-machine/
- Noble, Safiya U. *Algorithms of Oppression: How Search Engines Reinforce Racism.* (2018)
- Noble, Safiya U., and Tynes, Brendesha (Eds.). "The Intersectional Internet: Race, Sex, and Culture Online." *Peter Lang: Digital Formations Series*. NY (2016).
- Noble, Safiya U., and Roberts, Sarah T. "Through Google-Colored Glass(es): Design, Emotion, Class, and Wearables as Commodity and Control." In S. Tettegah & S. Noble (Eds.), *Emotions, Technology & Design*. pp. 187-210. San Diego: Elsevier Academic Press (2016).
- Noble, Safiya U. *Trayvon, race, media and the politics of spectacle*. The Black Scholar. 44(1), 12-29 (2014).
- Noble, Safiya U. *Google search: Hyper-visibility as a means of rendering black women and girls invisible*. InVisible Culture: Issue 19 (2013).
- O'Neil, Cathy. *Weapons of Math Destruction: How Big Data increases Inequality and Threatens Democracy*. New York: Broadway Books (2016).
- Paris, Britt, and Donovan, Joan. *Deepfakes and Cheap Fakes*. New York, NY: Data & Society Research Institute (September 18, 2019). datasociety.net/output/deepfakes-and-cheap-fakes/.
- Phillips, Jonathan P., and Jiang, Fang. *An other-race effect for face recognition algorithms*. www.semanticscholar.org/paper/An-other-race-effect-for-face-recognition-Phillips-Jiang/b995633ff8732110ff8e34fc8f04699dee84bdf5#citing-papers.

- powell, john a., and Menendian, Stephen. "The Problem of Othering: Towards Inclusiveness and Belonging." *Othering and Belonging* no. 3 (June 29, 2017). www.otheringandbelonging.org/the-problem-of-othering/
- Responsible Data. *Open Letter to WFP re: Palantir Agreement*. 2–8 (2019). Retrieved from responsibledata.io/2019/02/08/open-letter-to-wfp-re-palantir-agreement/.
- Roberts, Sarah T. *Digital Refuse: Canadian Garbage, Commercial Content Moderation and the Global Circulation of Social Media's Waste*, Wi: Journal of Mobile Media, 10(1) (2016): p 1-18.
- Roberts, Sarah T. and Noble, Safiya U. *Empowered to name, inspired to act: Social responsibility and diversity as calls to action in the LIS context*. Library Trends, 64(3) (2016): p. 512-532.
- Roberts, Sarah T. *In/visibility*. In Letters & Handshakes (Eds.), Surplus3: Labour and the Digital. Toronto: Letters & Handshakes (2016).
- Roberts, Sarah T. "Commercial content moderation: Digital laborers' dirty work." In Noble, S.U. and Tynes, B. (Eds.), *The intersectional internet: Race, sex, class and culture online*. New York: Peter Lang (2016).
- Sim, Kate, and Cheesman, Margie. *What's the harm in categoriazation? Reflections on the categorization work of Tech 4 Good*, Big Data & Society blog (2020).
- Smith, Carly Parnitzke, and Freyd, Jennifer J. *Institutional Betrayal*. American Psychologist 69, no. 6 (2014): 575–87. doi.org/10.1037/a0037564.
- Van der Kolk, Bessel A. *The Body Keeps the Score: Brian, Mind, and Body in the Healing of Trauma*. New York: Penguin Books (2015).
- Yap, Audrey S. *Credibility Excess and the Social Imaginary in Cases of Sexual Assault*. Feminist Philosophy Quarterly 3, no. 4 (2017): 1–24. doi.org/10.5206/fpq/2017.4.1.
- Yeung, Bernice. *In a Day's Work*. First edition. New Press (2018). thenewpress.com/books/days-work.

The Promise Institute for Human Rights
UCLA School of Law

www.law.ucla.edu/promiseinstitute

@PromiseInstUCLA

UCLA Center for Critical Internet Inquiry

www.c2i2.ucla.edu

@C2i2_UCLA

WITH SUPPORT FROM THE KNIGHT FOUNDATION

KNIGHT FOUNDATION