NEW AMERICA

October 2020

Automated Intrusion, Systemic Discrimination

How Untethered Algorithms Harm Privacy and Civil Rights

Christine Bannan & Margerite Blase

Acknowledgments

The authors would like to thank Representative Yvette Clarke, A. Prince Albert III, Koustubh "K.J." Bagchi, Daniel Kahn Gillmor, and Iris Palmer for participating in the event, and Austin Adams, Lisa Johnson, and Joe Wilkes for communications support. We would also like to thank Craig Newmark Philanthropies for its generous support of this work. The views expressed in this report are those of its authors and do not necessarily represent the views of Craig Newmark Philanthropies, its officers, or its employees.

About the Author(s)

Christine Bannan is policy counsel at New America's Open Technology Institute, focusing on platform accountability and privacy.

Margerite Blase is a Legal/Public Policy intern with New America's Open Technology Institute, working with the platform accountability and privacy teams.

About New America

We are dedicated to renewing the promise of America by continuing the quest to realize our nation's highest ideals, honestly confronting the challenges caused by rapid technological and social change, and seizing the opportunities those changes create.

About Open Technology Institute

OTI works at the intersection of technology and policy to ensure that every community has equitable access to digital technology and its benefits. We promote universal access to communications technologies that are both open and secure, using a multidisciplinary approach that brings together advocates, researchers, organizers, and innovators.

Contents

Introduction

Machine learning and other artificial intelligence (AI) tools are increasingly used by organizations and online platforms to help with critical, life-altering decisions. These include deciding whether, or for how long, someone should go to jail, whether someone should be considered for an open job, whether someone is likely to succeed at a university, and more. Because these systems rely on large datasets and statistical analyses, their outputs are often perceived as neutral and not affected by biases in the same way as human decision-making. However, outputs from algorithms and other automated tools can and do reinforce biases and lead to disparate results because the datasets they rely on reflect historical and existing discrimination. AI often requires a significant volume of personal data to function—collecting these data can require privacy-intrusive practices that also violate users' civil rights. The government and other stakeholders must acknowledge the risks posed by algorithmic decision-making and other automated tools to help protect those most likely to be harmed by them and to ensure these tools are used in non-discriminatory and beneficial ways. Last year, to address these issues, Rep. Yvette D. Clarke (D-N.Y.), along with Sens. Cory Booker (D-N.J.) and Ron Wyden (D-Ore.), introduced the Algorithmic Accountability Act of 2019.[1] The bill requires companies to test and fix flawed computer algorithms that result in biased or discriminatory outcomes.

This report builds on an **event** hosted on June 3, 2020 by New America's Open Technology Institute (OTI) that explored how machine learning and other algorithmic tools can lead to privacy and civil rights harms.[2] Rep. Clarke, the vice chair of the Energy and Commerce Committee, delivered opening remarks. A subsequent panel, moderated by Koustubh "K.J." Bagchi, OTI senior policy counsel, brought together Daniel Kahn Gillmor, senior staff technologist of the ACLU Project on Speech, Privacy, and Technology; Iris Palmer, senior advisor for higher education and workforce at New America's Education Policy program; and A. Prince Albert III, then-technology & telecommunications fellow at the Leadership Conference on Civil and Human Rights. The panel discussed key questions regarding algorithms and their potential impacts on privacy and civil rights, including: How do we design and audit algorithms to avoid disparate outcomes? What are the real-world consequences of algorithmic practices? Who should be held accountable to protect individuals from the discriminatory impacts of automated systems? What should legislative protections look like? To what extent can, or should, such protections be incorporated into comprehensive consumer privacy legislation?

Editorial disclosure: This report discusses policies by Facebook and Google, both of which are funders of work at New America but did not contribute funds directly to the research or writing of this report. New America is guided by the principles of full transparency, independence, and accessibility in all its activities and partnerships. New America does not engage in research or educational activities directed or influenced in any way by financial supporters. View our full list of donors at ***www.newamerica.org/our-funding****.*

Algorithmic Tools Used in Criminal Justice, Education, and Employment Are Powered by Personal Data

Machine learning and other algorithmic tools are increasingly used to replace or augment human decision-making. A 2019 survey conducted by New Vantage Partners showed that 91.6 percent of the Fortune 1000 executives surveyed were increasing their big data and AI investments over the previous year.[3] Algorithms powered by vast quantities of personal data are now commonly used to make decisions about numerous aspects of people's lives. OTI's event and this report focus on criminal justice, education, and employment, but the widespread use of automated tools also affects many other sectors, including credit and housing. Both the data used to train algorithms and the data collected and used to create outputs present privacy and equity issues. The outputs from machine learning algorithms are created from the underlying data, the coded instructions, and the algorithms' own learnings. Even when their creators did not intend to discriminate against certain groups, algorithmic tools can reinforce historical discrimination if the training data used to build the AI reflects biases.[4] After the AI is trained, the tools can collect and store sensitive data—such as gender, race, and medical conditions—that can increase the scope of personal information used to make decisions for or about an individual.[5]

Algorithmic systems can harm individual privacy, both in the collection required to build the system and after the system is built. AI systems based on statistical models require large data sets to function, and there are privacy risks inherent in collecting this volume of data on individuals. Algorithms reliant on machine learning constantly require more data to train the system and enable it to draw inferences. When systems require such volumes of data about people, it is likely that this increases the risk of data being obtained in a privacy-intrusive manner.[6] During the panel, Gillmor noted that when an entity stores large data sets that it does not properly manage, this can become an "attractive nuisance" for law enforcement, immigration services, foreign hackers, or identity thieves to target.

Algorithmic Tools Used in Criminal Justice

Automated tools are widely used in the criminal justice system, and their use has generally led to inequitable outcomes.[7] Two of the most commonly used tools are risk assessment algorithms and facial recognition tools. Risk assessment algorithms are used to estimate the likelihood of certain outcomes, such as a defendant's chance of recidivism or failure to appear before a judge. As Albert III noted during the panel, risk assessment "can be used at really any juncture of the criminal legal process where critical decisions are made of freedom." They are commonly used at the pre-trial stage[8] to replace or supplement cash bail systems, [9] during incarceration to make determinations about early release, and post-release to make decisions about probation and parole. These tools use information such as criminal history, socioeconomic status, neighborhood crime rates, and other factors to predict an individual's potential risks, and can be used without an individual's consent. When used to predict future criminal behavior, these tools raise grave civil liberties concerns.

OTI joined a coalition of over 100 civil rights, digital justice, and community-based organizations in a statement opposing these types of tools in the criminal justice system and condemning the use of algorithmic assessments for pretrial detention because risk assessment tools can result in racially biased outcomes that reflect patterns of historical discrimination.[10] The statement also called for safeguards and testing requirements in cases where such tools are already in use. Black people are one-third more likely to be stopped by the police and three times more likely to be searched by the police.[11] Algorithms that rely on police records and criminal history to predict likelihood of recidivism will perpetuate these discriminatory patterns found in the training data, and disproportionately harm Black people.[12] Yet decisions from these automated decision-making systems often do not face the same amount of scrutiny as their human counterparts due to an assumption that technical solutions are inherently more accurate and objective.

The First Step Act, signed into law in 2018, requires the U.S. Attorney General to develop a "risk and needs assessment system" for the Federal Bureau of Prisons to assess each prisoner's risk of recidivism and determine what type of recidivism reduction programming is appropriate for them.[13] To address this law, the U.S. Attorney General William P. Barr and the United States Department of Justice (DOJ) created the Prisoner Assessment Tool Targeting Estimated Risk and Need (PATTERN). As of January 2020, the DOJ announced that incarcerated individuals will be assigned into recidivism reduction programs and other activities based on their assessment scores from PATTERN.[14] Those individuals who participate or complete these programs can be placed in pre-release custody or receive sentence reductions, indicating how important scores from the risk assessment tools can be for prisoners. Another example of this occurred earlier this year when the Barr ordered that some federal prisoners should be released to avoid overcrowding during the COVID-19 pandemic, and that prisoners with a minimum PATTERN score should be prioritized. This requirement can favor white prisoners over Black prisoners, who are more likely to have higher risk assessment scores due to historical patterns of disparate policing practices.[15] As Albert III noted, these risk assessment scores can exacerbate racial inequalities and should not be used to "essentially prioritize who lives and who dies" during a pandemic.

Algorithmic Tools Used in Education

Educational institutions at all levels use algorithms to make critical decisions for students, such as which curriculum they study or what resources they should receive. The different ways algorithms are used can either perpetuate discrimination or help address it. One example of this is through ability grouping, where schools divide students into different groups based on their academic ability. This practice has been used for decades, but recent changes in technology allow educational data-mining (EDM) technologies to sort through vast amounts of student data to form these groupings.[16] It is particularly important to understand the repercussions of using these types of algorithms in education. These systems can determine what skills students develop, the curriculum they are taught, who their peers are, and what expectations teachers have of them.

EDM technologies can, however, have a positive influence on the education system as well. As Palmer explained, when these systems are well-designed, they can help higher education systems "reach out to the students that are vulnerable, who need their help, and who would not have otherwise come through (an advisor's) door." An experiment at Georgia State University found that students who received assistance through an AI outreach tool that provided guidance on the application process were more likely to enroll than their control group counterparts.[17] This type of assistance can have a particularly positive effect on low-income and first-generation students who may not otherwise have outside guidance on the college application process.[18]

However, similar to the issues caused by algorithmic systems in criminal justice, education algorithms learn to make predictions from training datasets that include historical data and can therefore reinforce patterns of racial, gender, and socioeconomic discrimination. Schools use predictive analytics tools that have been trained on the data of past students to select which attributes best predict a student's success and forecast whether a student will perform well. As Palmer noted, this can contribute to the underrepresentation of minority students in science, technology, engineering, and mathematics (STEM) fields. When algorithms predict that students will not be successful in a certain major, students could be discouraged from pursuing that field of study. When schools use predictive analytics tools, they are relying heavily on the data available, which reflect inequities in the education system and therefore may not accurately assess current students' abilities. Students from a privileged background are more likely to have access to technology and be more technologically proficient compared to other students. This means that algorithms that rely on the amount of time students spend logged into an educational resource, for instance, may not accurately capture the academic ability or learning habits of a student from a low-income background if the student lacks internet access or otherwise has inadequate access to the resource.[19]

The use of algorithms in the education system can also pose privacy and security risks. Schools and universities collect extensive personal data on students. This can include what courses a student has taken and the clubs they are involved in, as well as a student's home address or medical history.[20] Although there is a federal law protecting student privacy, the Family Educational Rights and Privacy Act (FERPA), it has a limited scope and has not kept pace with the dramatic changes in educational technology and student data collection.[21] Schools often partner with different third-party vendors that help them run their online education programs, supply predictive analytics services, and provide other technology and data support. But under FERPA, only schools are responsible for what vendors do with that data.[22] Both schools and vendors often lack sufficient protocols to protect student data, causing extensive security and privacy concerns. Sophisticated cyber attacks are not necessarily the most likely reason for a breach, particularly at universities, where Palmer explained that most breaches are due to the system lacking basic security practices. Since 2005, K-12 school districts, colleges, and universities in the United States have experienced over 1,300 data breaches affecting more than 24.5 million student records.[23]

Algorithmic Tools Used in Employment

A growing number of organizations are also utilizing machine learning algorithms to make employment decisions, such as who they should interview or hire. This is another important area where the use of algorithms, without proper testing and assessment tools, can perpetuate historical biases and negatively affect certain minority groups. For example, if a recruiting system makes decisions on candidates based on predicted tenure at an employer, the system may be more likely to have a disparate impact on certain protected classes. As Gillmor explained during the panel, if a system is built to simply compare candidates to people who have already done well in an organization "and the reason some people are not doing well at a company is an internally discriminatory regime, a system will pick up on that" and assess potential new candidates based on this regime, thereby reenforcing discrimination.

When employers pay for online job advertisements, internet platforms use machine learning algorithms to both target and distribute the ads.[24] Facebook, like other advertising platforms, allows advertisers to select and target audiences based on demographic factors. Until 2019, the company allowed advertisers to select or exclude users from being shown advertisements based on protected characteristics, such as gender and race,[25] a practice they ended in response to lawsuits pertaining to housing, employment, and credit ads that ran on their website. Although Facebook has since removed the option to target or exclude protected classes from certain advertisements, studies have shown that their advertising algorithm may still result in racial or gender biases in the delivery of ads. Because training data for these algorithms reflect historical employment discrimination, employers who do not intend to only target certain demographics may still have their ads delivered primarily to people of a certain race or gender. For example, researchers that ran five ads for jobs in the lumber industry that tried to deliver them to a large and inclusive audience found that the ad delivery algorithm delivered to over 90 percent male users and over 70 percent white users in aggregate.[26] Researchers also found that the delivery of housing and employment ads on Facebook was skewed based only on the ad's content and link.[27] Researchers have also found Google's advertising algorithm perpetuated biases in employment advertising. In one study, the researchers posed as male and female users, and found that Google served ads for high-paying executive positions at a higher rate to the male profiles.[28] Though it is hard to confirm exactly why these types of issues are occurring without further insight into the proprietary algorithmic tools, the results of these studies show that the advertising systems are perpetuating patterns of discrimination found in the training data.

Lack of Transparency and Perceived Objectivity Perpetuate Biases in Algorithmic Tools

As Gillmor noted, machine learning and other AI systems are primarily used as cost-shifting measures. He argues, they "are designed to make it possible to do things at a scale that your traditional mechanisms would not be able to do." Companies that adopt AI technologies report having an increase in revenue, as well as reduced costs.[29] These technologies can also benefit society by predicting natural disasters or improving medical diagnostics, for example.[30] Algorithmic tools will be more appropriate for uses that do not rely on sensitive personal information and do not make consequential decisions about individuals. However, there is often a lack of transparency and understanding of how these models work, making it difficult to judge their accuracy and social fairness. As more companies rely on and invest in big data and AI systems, it is important to ensure that these systems do not result in disproportionately negative outcomes, particularly for historically disadvantaged groups.

If AI is not designed and monitored properly, the technology can have discriminatory consequences. For example, while these systems can help improve medical diagnostics, a study on a widely used health care algorithm showed that they can also systematically discriminate against Black people.[31] The study found that the system failed to identify Black patients at risk for medical needs at the same rate as white patients, which resulted in Black patients being less likely to receive preventative care to improve their health. These results are particularly alarming during the COVID-19 pandemic, because Black and Brown communities are already disproportionately affected by the pandemic[32] and AI systems may be used to help determine how to prioritize medical care if resources are scarce.[33]

As Albert III noted, decision-makers feel comfortable relying on algorithmic systems because there is often a perception that automated systems are less biased than human decision-makers because they rely solely on data and statistics. When companies and organizations actually test AI systems, however, they often find unintended, biased results.

There are a variety of reasons why algorithms can lead to biased results. One of the most common reasons is the underlying training data, which reflect biases and discriminiation that exist in the physical world. Discrimination continues to persist in the American education and criminal justice systems, as well as among employers in all sectors of the economy. Because many institutions were built on a foundation of racism and inequality, the data collected by those institutions will reflect those biases because predictive algorithms are designed based on correlations found in data of past experiences. So, as noted above, if historical data show that in the past only white men have held certain jobs, or that minorities who lack access to educational resources will not perform as well academically, these historical discriminatory patterns will be perpetuated by the algorithms. Further, discrimination can occur even when the underlying dataset does not explicitly contain sensitive categories such as race or gender, but only includes highly correlated variables that can serve as proxies for these characteristics.

For example, in 2018 Amazon built out and tested an AI recruiting tool to help find and review job applicants.[34] Over time, Amazon found that the tool had a preference for male candidates over female candidates for technical positions. The tool was not built to consider gender for potential applicants, but it was built to identify candidates based on patterns in previous resumes submitted to Amazon—historically, applications for technical positions had predominantly come from men. Amazon abandoned the tool before it was used to actually recruit potential candidates, but the unintended outcome shows the importance of testing and auditing AI systems for discriminatory impacts before and after deployment.

It can be particularly difficult to hold companies accountable for discrimination caused by their AI tools, because those affected often lack knowledge of how the tools work and the ability to obtain that information. The final rule released by the Department of Housing and Urban Development (HUD) in September 2020 will exacerbate this problem for those trying to hold housing providers responsible for discrimination, because it places so much evidentiary burden on the plaintiff. HUD's disparate impact rule, both in its proposed form (in 2019) and its final form (2020), show that HUD fails to recognize AI tools' likelihood of leading to discriminatory housing decisions. OTI filed comments as part of a coalition of 23 civil rights and consumer advocacy organizations and individual experts,[35] and separately,[36] to oppose the HUD proposal, explaining that it reflected a complete lack of understanding of how algorithmic models work. Although HUD removed the controversial algorithmic defense, the final rule essentially has the same effect by significantly heightening the burden on plaintiffs and creating a new defense that permits disparate impact caused by an algorithm if the defendant establishes the algorithm provided some benefit to the protected class. In cases involving AI tools, plaintiffs must be more reliant on a showing of disparate impact, but HUD's rule will make it extraordinarily difficult for their claims to prevail. The new requirement that plaintiffs establish a "robust causal link between the policy or practice and the adverse effect on members of a protected class" make it particularly difficult for plaintiffs to bring a claim when the discriminatory impact is caused by AI, because they will likely not have access to enough information about internal, technical practices to satisfy this requirement. That information is always difficult for plaintiffs to obtain, as companies will claim "trade secret" protections, making algorithms extraordinarily difficult to challenge in court generally—but could be an impassable hurdle at the pleading stage (as the new rule requires).[37]

Similarly, a *ProPublica* study showed that the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), a criminal risk assessment tool, is twice as likely to classify Black defendants who do not recidivate as high risk compared to similar white defendants.[38] Based on these findings, *ProPublica* found that COMPAS is unreliable and racially biased against Black defendants. Northpointe, the developer of COMPAS, asserted that the system is not racially biased because it predicts overall recidivism equally well for Black and white defendants. The discrepancy appears to stem from the fact that there have historically been greater levels of policing in Black communities, and consequently disproportionately higher arrest rates.[39] However, it is difficult to confirm these findings because the COMPAS algorithm is proprietary data that Northpointe has not released.[40] In *State v. Loomis*, the defendant argued that the use of COMPAS violated his due process rights because the tool considers gender in its analysis, although he could not confirm the specific method in which it did so because the algorithm was proprietary.[41] The Wisconsin Supreme Court upheld the use of COMPAS, but required that COMPAS reports be accompanied by disclaimers of accuracy. Without accurate knowledge of how the COMPAS's algorithm works, however, it will be difficult for any future defendants who suspect they were unfairly discriminated against to oppose their risk assessment score. Because of the evidence of discrimination in risk assessment tools, OTI and over 100 organizations have called on jurisdictions across the United States to abandon their use in pretrial decisionmaking.[42]

Further, AI systems are often not equipped to quickly adapt to unprecedented times. As Palmer noted, the pandemic's effects on higher education have caused student data to differ from expectations before the pandemic, which can decrease the accuracy of a tool's predictions dramatically. For example, universities often use algorithms to identify students who may be at risk of dropping out or failing. One of the metrics these tools consider is how often a student is engaging with an online learning management system. However, during the pandemic, many classes and other university services have gone remote, so all students use online-learning management systems to a higher degree. Students who were previously flagged as being at-risk are not anymore because their online engagement went up, although, according to Palmer, "all students are more at risk in this situation." Because of the pandemic, these systems will no longer be as effective at identifying students who need more help, and universities will need to consider additional methods for identifying and helping students. There must be proper assessment systems in place to ensure that unprecedented shifts in data do not render algorithm decision-making ineffective.

AI can be extremely valuable to society and help organizations accomplish tasks that would otherwise not be possible. However, without proper testing and assessment tools in place, it can also lead to unintentional discriminatory outcomes. Before AI and algorithmic systems are used for autonomous decision-making, there should be more processes in place for auditing and assessing these tools to ensure their outcomes are not biased.

Recommendations for Preventing Abusive Algorithmic Tools

Addressing the potentially harmful outcomes of AI will require a critical examination of these tools at each stage of their development and implementation. Algorithmic tools disproportionately harm people of color—especially Black and Brown communities—women, immigrants, religious minorities, members of the LGBTQ+ community, low-income individuals, and other marginalized communities. In addition to discriminatory impacts, the expansive collection of personal data necessary to power the algorithmic tools used in criminal justice, education, and employment settings harms individual privacy. Therefore, it is essential to center civil rights in the privacy debate and advocate for policies that would prevent discriminatory data practices.[43] Transparency requirements, impact assessments, and regular audits are all necessary and can be mandated in comprehensive privacy legislation. Whether or not there are legal requirements, institutions need to evaluate the consequences of algorithmic tools and determine if it is possible to deploy these systems equitably.

Transparency is a necessary, but not sufficient, condition for reforming discriminatory algorithmic systems. According to Gillmor, "the more that we can understand these systems the better we can defend ourselves against them." Some systems use less complex algorithms that are more easily explained. Albert III noted that many risk assessments are more accurately described as calculators than as AI, and simply project the generalized assumptions about recidivism of demographic groups onto an individual belonging to that group. For example, the age of a defendant is a heavily weighted demographic in risk assessments. The system used in Virginia assigns a defendant more risk points for being 23 years old (as opposed to being older) than for having five or more prior incarcerations.[44] But, Albert III explained, if criminal defendants do not know what factors are used to calculate a risk assessment, they will not be able to defend themselves adequately. Therefore if a pretrial risk assessment is used, it "must be transparent, independently validated, and open to challenge by an accused person's counsel."[45] For risk assessments, the current lack of transparency is often a legal issue rather than a technical issue.

When algorithms are more complex, such as systems based on machine learning, transparency can be a technical challenge. For black box systems, even the developers who created the system may be unable to explain precisely how the system operates and why it produced a particular output. Yet, when such systems are used to make critical decisions about a person's life—including criminal justice, education, and employment determinations—it is not acceptable for the algorithms to be opaque and sealed from scrutiny. It is particularly important for these types of algorithms to undergo intensive audits.

Before an institution decides to implement an AI tool, it should conduct an impact assessment to evaluate the potential risks. Palmer recommends that colleges require a predictive analytics vendor to agree to conduct a disparate impact analysis before entering a contract. To assist education institutions, she wrote a guide that colleges can use before partnering with a vendor.[46] The same principles apply to other institutions that hire vendors to create algorithmic models to predict outcomes for individuals; they should provide transparency, privacy, security, equity, and regular evaluations.

After an AI system is deployed, it should also undergo regular audits to detect flaws or harmful consequences. Since software requires continuous updates to fix bugs and make improvements, audits need to be ongoing to stay up to date with the current version of the software. Moreover, when a system introduces new training data, audits must be updated to include an evaluation of that new data. Although audits can be costly, due process does not become less important because it is resource intensive. If an institution finds that regular audits will be too costly, they should reconsider using AI systems. Organizations and entities who perform audits need to consider the possibility that a system should be temporarily or permanently suspended if a discriminatory impact or other harmful results are identified. For example, at the urging of Facebook employees, in July 2020 the company launched an investigation into whether its machine learning algorithms discriminate against minority ethnic groups.[47] This internal review should be a helpful step in light of a final audit report, conducted by the company at the urging of civil society groups, which outlined a variety of discriminatory practices and made a series of concrete recommendations to address the harms the company perpetuates.[48] But in order for the investigation to be meaningful, it must include the possibility that Facebook will need to suspend a profitable business practice.

Some AI systems could be used both to perpetuate historical discrimination or to attempt to confront it. Criminal justice authorities, institutions of higher education, and employers should evaluate their reasons for using an algorithmic tool and how they are framing the questions they want the tool to answer. People are prone to misinterpreting probabilities,[49] and conclusions based upon faulty assumptions about the meaning of an algorithmic model's predictions can have detrimental consequences. Palmer recommends that universities train advisors how to interpret and convey predictive data to avoid misuse of the algorithm's outputs. For example, instead of a college advisor using a predictive model to discourage a student from pursuing a certain major, that advisor could use the same output to direct the student toward appropriate resources. In the criminal justice context, the way a risk assessment is presented to a judge or jury is essential to preserving the presumption of innocence. If risk assessments are used at all, they should only identify groups of people to be released immediately because automated recommendations for detention assume the guilt of the defendant.[50]

At the most fundamental level, institutions need to consider reframing the questions they ask these tools to answer. Algorithmic systems that reflect biases within an institution can be used to reform that institution rather than to make discriminatory decisions about individuals. As Gillmor noted, when a hiring system categorizes women as less likely to succeed at a company than men, that finding should be used to address a corporate culture of sexism rather than for hiring. Amazon correctly decided not to use its system in hiring after detecting that it was reflecting such biases. The next step would be for the company to examine why its system found that women were less likely to succeed. Palmer explained that risk assessments that identify Black students as less likely to graduate are reflective of systemic patterns in higher education, and universities should examine how they are failing to support those students. However, in some instances, it is better for institutions to stop using an algorithmic system rather than attempt to use it to address bias. In Albert III's opinion, due to the amount of structural inequity in the criminal justice system, there is currently no way that risk assessment tools can be used equitably.

Legislation that requires transparency measures, impact assessments, and audits would help prevent abusive uses of algorithmic tools and mitigate discriminatory harms. Rep. Clarke's bill requires entities using automated decision systems to conduct impact assessments and mandates additional safeguards for "high-risk information systems" such as those that use data about sensitive characteristics including race, gender, biometrics, and criminal arrests.[51] The assessments must analyze characteristics that are central to traditional privacy legislation, including: data minimization practices, the retention period for personal information, and the ability of consumers to access and object to or correct the results of the automated decision system. Congress could pass these requirements as standalone legislation, or incorporate them into privacy legislation.

Another legislative tool to address the concerns raised is comprehensive privacy legislation. Comprehensive privacy legislation should require transparency, impact assessments, and regular audits to prevent algorithmic tools from being used in ways that disparately impact disadvantaged communities. Currently, U.S. law generally relies on "notice and consent" to protect consumer privacy, but this framework does not give individuals real choices about how their data are used and is insufficient to protect user privacy.[52] There is a growing consensus among stakeholders to abandon this model in favor of a new approach that places restrictions on how data can be used and gives users enforceable rights over their personal information. In 2018, OTI, as part of a group of 34 civil rights, consumer, and privacy organizations, released public interest principles for privacy legislation, including the principle that "[d]ata practices must protect civil rights, prevent unlawful discrimination, and advance equal opportunity."[53]

Since then, members of Congress have introduced bills that include specific requirements for algorithmic tools. U.S. Senate Committee on Commerce, Science, and Transportation Ranking Member Maria Cantwell (D-Wash.) introduced the Consumer Online Privacy Rights Act that would require entities that advertise or make eligibility determinations for housing, education, employment, or credit opportunities to conduct "algorithmic decision-making impact assessments."[54] The assessments must describe and evaluate the design of an algorithmic tool and the training data used and determine whether the decision-making system produces discriminatory results. The most recent comprehensive privacy bill, Sen. Sherrod Brown (D-Ohio)'s Data Accountability and Transparency Act of 2020, would go a step further and require "automated decision system risk assessments" to be made publicly available.[55] The growing recognition of privacy as a civil right in Congress is a promising signal that the United States can address the risks posed by the increasing use of algorithmic tools.

Conclusion

Algorithmic tools used to make consequential decisions about individuals' lives are becoming increasingly common in criminal justice, higher education, and employment, as well as other sectors. Too often, these AI systems are trained with data sets reflecting historical biases and collected through privacy-invasive means. A multi-level approach is required to prevent discriminatory, privacy-invasive, and other harmful outcomes. Such systems must be transparent, undergo a disparate impact analysis before implementation, and undergo regular audits after implementation. Most fundamentally, equitable use of AI systems requires institutions to examine their own biases and honestly assess whether the benefits of a tool could outweigh the risks to privacy and equity.

Notes

[1] Algorithmic Accountability Act, H.R. 2231, 116th Congress (2019), https://www.congress.gov/bill/116th-congress/house-bill/2231.

[2] Rep. Yvette Clarke, A. Prince Albert III,Daniel Kahn Gillmor, Iris Palmer, Koustubh Bagchi, "Automated Intrusion, Systemic Discrimination," (Panel, Online, June 3, 2020), https://www.newamerica.org/oti/events/online-automated-intrusion-systemic-discrimination/

[3] "Big Data and AI Executive Survey 2019," NewVantage Partners LLC (2019), https://newvantage.com/wp-content/uploads/2018/12/Big-Data-Executive-Survey-2019-Findings-Updated-010219-1.pdf

[4] See e.g., Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, Cass R Sunstein, "Discrimination in the Age of Algorithms," *Journal of Legal Analysis,* 10, (April 2019), https://doi.org/10.1093/jla/laz001

[5] Cameron F. Kerry, "Protecting Privacy in an AI-driven World," Brookings, February 10, 2020, https://www.brookings.edu/research/protecting-privacy-in-an-ai-driven-world/

[6] "Royal Free - Google DeepMind Trial Failed to Comply with Data Protection Law," Information Commissioner's Office, July 3, 2017, https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/07/royal-free-google-deepmind-trial-failed-to-comply-with-data-protection-law

[7] See e.g. "Algorithms in the Criminal Justice System: Risk Assessment Tools," EPIC, accessed August 10, 2020, https://epic.org/algorithmic-transparency/crim-justice

[8] Electronic Privacy Information Center. "Liberty At Risk: Pre-Trial Risk Assessment Tools in the U.S." epic.org, July 2020. https://epic.org/LibertyAtRiskReport.pdf.

[9] Doyle, Colin, Chiraag Bains, and Brook Hopkins. "Bail Reform: A Guide for State and Local Policymakers." Criminal Justice Policy Program. Harvard Law School, February 2019. http://cjpp.law.harvard.edu/assets/BailReform_WEB.pdf.

[10] See e.g. "New America's Open Technology Institute Joins Coalition Condemning Use of Algorithmic Risk Assessments for Pretrial Detention," New America, July 30, 2018, https://www.newamerica.org/oti/press-releases/new-americas-open-technology-institute-joins-coalition-condemning-use-algorithmic-risk-assessments-pretrial-detention/

[11] Christine Kumar, "The Automated Tipster: How Implicit Bias Turns Suspicion Algorithms into BBQ Beckys", 72 Fed. Comm. L.J. 97, (June 2020), http://www.fclj.org/wp-content/uploads/2020/06/Volume72.1-ChristineKumar.pdf

[12] See e.g. Greg Satell, Josh Sutton, "We Need AI That Is Explainable, Auditable, and Transparent," Harvard Business Review, October 28, 2019, https://hbr.org/2019/10/we-need-ai-that-is-explainable-auditable-and-transparent

[13] First Step Act of 2018, S. 756 (115th Cong.), https://www.congress.gov/bill/115th-congress/senate-bill/756/text

[14] "Department of Justice Announces Enhancements to the Risk Assessment System and Updates on First Step Act Implementation," Department of Justice, January 15, 2020, https://www.justice.gov/opa/pr/department-justice-announces-enhancements-risk-assessment-system-and-updates-first-step-act

[15] See e.g. Nathan James, "Risk and Needs Assessment in the Criminal Justice System," Congressional Research Service, (October 2015), https://digital.library.unt.edu/ark:/67531/metadc795663/

[16] Yoni Har Carmel, Tammy Harel Ben-Shahar, "Reshaping Ability Grouping Through Big Data, " Vanderbilt Journal of Entertainment & Technology Law, (May 2017), https://ssrn.com/abstract=2944743

[17] Lindsey C. Page, Hunter Gehlbach, “How an Artificially Intelligent Virtual Assistant Helps Students Navigate the Road to College,” AERA Open, December 12, 2017, https://journals.sagepub.com/doi/10.1177/2332858417749220#articleCitationDownloadContainer

[18] See e.g. Laura Falcon, “Breaking Down Barriers: First-Generation College Students and College Success,” League for Innovation in the Community College, June, 2015, https://www.league.org/innovation-showcase/breaking-down-barriers-first-generation-college-students-and-college-success

[19] Closing the Home Learning and Homework Gap: Innovative School and Community Wi-Fi Initiatives, New America’s Open Technology Institute, June 25, 2020, https://www.newamerica.org/oti/events/closing-home-learning-and-homework-gap/.

[20] Jonah Newman, “Do you know what your college is doing with your data?,” Marketplace, September 25, 2014, https://www.marketplace.org/2014/09/25/do-you-know-what-your-college-doing-your-data/

[21] “Legislative History of Major FERPA Provisions,” U.S. Department of Education, accessed August 14, 2020, https://www2.ed.gov/policy/gen/guid/fpco/ferpa/leg-history.html

[22] Tina Nazerian, “The Unintentional Ways Schools Might Be Violating FERPA, and How They Can Stay Vigilant,” EdSurge, September 12, 2018, https://www.edsurge.com/news/2018-09-12-the-unintentional-ways-schools-might-be-violating-ferpa-and-how-they-can-stay-vigilant

[23] Cook, Sam. “US Schools Leaked 24.5 Million Records in 1,327 Data Breaches since 2005.” Comparitech, July 1, 2020. https://www.comparitech.com/blog/vpn-privacy/us-schools-data-breaches/.

[24] Spandana Singh, Special Delivery: How Internet Platforms Use Artificial Intelligence to Target and Deliver Ads, New America’s Open Technology Institute, February 18, 2020, https://www.newamerica.org/oti/reports/special-delivery/

[25] Colin Lecher, “Facebook drops targeting options for housing, job, and credit ads after controversy,” The Verge, March 19, 2019, https://www.theverge.com/2019/3/19/18273018/facebook-housing-ads-jobs-discrimination-settlement

[26] Muhammad Ali et al., "Discrimination through optimization: How Facebook’s ad delivery can lead to skewed outcomes," Arxiv, September 12, 2019, https://arxiv.org/pdf/1904.02095.pdf

[27] Muhammad Ali et al., "Discrimination through optimization: How Facebook’s ad delivery can lead to skewed outcomes," Arxiv, September 12, 2019, https://arxiv.org/pdf/1904.02095.pdf

[28] Amit Datta, Michael Carl Tschantz, and Anupam Datta, Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination, April 18, 2015, Proceedings on Privacy Enhancing Technologies, https://doi.org/10.1515/popets-2015-0007

[29] “Global AI Survey: AI Proves its Worth, but Few Scale Impact,” McKinsey, November 22, 2019, https://www.mckinsey.com/featured-insights/artificial-intelligence/global-ai-survey-ai-proves-its-worth-but-few-scale-impact

[30] See e.g. Naveen Joshi, “How AI Can And Will Predict Disasters,” Forbes, March 15, 2019, https://www.forbes.com/sites/cognitiveworld/2019/03/15/how-ai-can-and-will-predict-disasters/#48ec02255be2 ; Darrell M. West, John R. Allen, “How artificial intelligence is transforming the world,” Brookings, April 24, 2018, https://

www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/

[31] “There is No Such Thing as Race in Health-care Algorithms,” The Lancet Digital Health, (December, 2019), https://www.thelancet.com/journals/landig/article/PIIS2589-7500(19)30201-8/fulltext

[32] CDC, “Health Equity Considerations and Racial and Ethnic Minority Groups,” July 24, 2020, https://www.cdc.gov/coronavirus/2019-ncov/community/health-equity/race-ethnicity.html

[33] Alex Engler, “A guide to healthy skepticism of artificial intelligence and coronavirus,” Brookings, April 2, 2020, https://www.brookings.edu/research/a-guide-to-healthy-skepticism-of-artificial-intelligence-and-coronavirus/

[34] Jeffrey Dastin, “Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women,” Reuters, October 9, 2018, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

[35] “OTI Joins Coalition Opposing HUD-Proposed Changes that Would Undermine Fair Housing Act Enforcement,” New America, October 18, 2019, https://www.newamerica.org/oti/press-releases/oti-joins-coalition-opposing-hud-proposed-changes-would-undermine-fair-housing-act-enforcement/

[36] New America’s Open Technology Institute’s Comment on HUD Proposed Rule, October 18, 2019, https://newamericadotorg.s3.amazonaws.com/documents/New_Americas_Open_Technology_Institute_Comments_on_HUD_Proposed_Rule_FR-6111-P-02.pdf

[37] John Villasenor and Virginia Foggo, “Why a Proposed HUD Rule Could Worsen Algorithm-Driven Housing Discrimination,” The Brookings Institution, April 16, 2020. https://www.brookings.edu/blog/techtank/2020/04/16/why-a-proposed-hud-rule-could-worsen-algorithm-driven-housing-discrimination/

[38] Jeff Larson, Surya Mattu, Lauren Kirchner, Julia Angwin, "How We Analyzed the COMPAS Recidivism Algorithm," ProPublica, May 23, 2016, https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

[39] See e.g., Alex Chohlas-Wood, “Understanding risk assessment instruments in criminal justice,” Brookings, June 19, 2020, https://www.brookings.edu/research/understanding-risk-assessment-instruments-in-criminal-justice/

[40] “Racial Bias and Gender Bias Examples in AI Systems.” Medium. The Comuzi Journal, (September, 2018), https://medium.com/thoughts-and-reflections/racial-bias-and-gender-bias-examples-in-ai-systems-7211e4c166a1

[41] State v. Loomis, 371 Wis. 2d 235 (2016)

[42] See e.g. “New America’s Open Technology Institute Joins Coalition Condemning Use of Algorithmic Risk Assessments for Pretrial Detention,” New America, July 30, 2018

[43] Chao, Becky, Eric Null, and Brandi Collins-Dexter. “Centering Civil Rights in the Privacy Debate.” New America's Open Technology Institute, August 14, 2019. https://www.newamerica.org/oti/reports/centering-civil-rights-privacy-debate/.

[44] Stevenson, Megan and Doleac, Jennifer L., Algorithmic Risk Assessment in the Hands of Humans (November 18, 2019). Available at SSRN: https://ssrn.com/abstract=3489440 or http://dx.doi.org/10.2139/ssrn.3489440

[45] “The Use of Pretrial ‘Risk Assessment’ Instruments: A Shared Statement of Civil Rights Concerns.” The Leadership Conference on Civil and Human Rights, July 30, 2018. http://civilrightsdocs.info/pdf/criminal-justice/Pretrial-Risk-Assessment-Full.pdf.

[46] Palmer, Iris. “Choosing a Predictive Analytics Vendor: A Guide for Colleges.” New America,

September 5, 2018. https://www.newamerica.org/education-policy/reports/choosing-predictive-analytics-vendor-guide/.

[47] Alex Hern, “Facebook to investigate claims its algorithms are discriminatory,” July 22, 2020, The Guardian, https://www.theguardian.com/technology/2020/jul/22/facebook-investigate-claims-algorithms-discriminate-ethnic-minorities

[48] Murphy, Laura W., and Megan Cacace. Rep. Facebook’s Civil Rights Audit – Final Report, July 8, 2020. https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf.

[49] Campbell, Don. New research uncovers why an increase in probability feels riskier than a decrease. University of Toronto Scarborough, June 20, 2016. https://ose.utsc.utoronto.ca/ose/story.php?id=8531%C2%A7id.

[50] “The Use of Pretrial ‘Risk Assessment’ Instruments: A Shared Statement of Civil Rights Concerns.” The Leadership Conference on Civil and Human Rights, July 30, 2018. http://civilrightsdocs.info/pdf/criminal-justice/Pretrial-Risk-Assessment-Full.pdf.

[51] Algorithmic Accountability Act, H.R. 2231, 116th Congress (2019), https://www.congress.gov/bill/116th-congress/house-bill/2231.

[52] Park, Claire. “How ‘Notice and Consent’ Fails to Protect Our Privacy.” New America. Open Technology Institute, March 23, 2020. https://www.newamerica.org/oti/blog/how-notice-and-consent-fails-to-protect-our-privacy/.

[53] “Principles for Privacy Legislation.” New America. Open Technology Institute, November 13, 2018. https://www.newamerica.org/oti/press-releases/principles-privacy-legislation/.

[54] Consumer Online Privacy Rights Act, S. 2968 (2019). https://www.congress.gov/bill/116th-congress/senate-bill/2968/text.

[55] Data Accountability and Transparency Act, discussion draft (2020), https://www.banking.senate.gov/imo/media/doc/Brown%20-%20DATA%202020%20Discussion%20Draft.pdf.