ALGORITHMIC ACCOUNTABILITY:

The Need for a New Approach to Transparency and Accountability When Government Functions Are Performed by Algorithms

By
Media Freedom & Information Access Clinic,
Yale Law School

JANUARY 18, 2022

Acknowledgements
This paper reflects the combined efforts of Yale Law School students Chloe Francis '23, David Froomkin '22, Eli Pales '22, Benjamin David Rashkovich '21, Karen Sung '23, and Kataeya Wooten '22, members of the Media Freedom & Information Access Clinic.

The Clinic is a program of the Abrams Institute for Freedom of Expression at Yale.

This discussion paper does not reflect the institutional views, if any, of Yale Law School or Yale University.

EXECUTIVE SUMMARY

Government authorities are increasingly using algorithms and machine learning technologies to conduct government business, allowing algorithms to make decisions on everything from assigning students to magnet schools, to allocating police resources, setting bail and distributing social welfare benefits. While algorithms promise to make government function more effectively, their growing use presents significant issues that policymakers have yet to address:

- Algorithms can make mistakes, either because they are poorly conceived or due to coding errors. Improperly functioning government algorithms can cause serious harm, such as by wrongly depriving people of health benefits or incorrectly identifying them as criminals.
- Algorithms can amplify pre-existing biases by being trained on biased historic data. Biased algorithms have been shown, for example, to increase the racially disparate impact of policing and cause the disproportionate removal of children from poorer parents.
- Algorithms are unaccountable. Agencies acquire algorithms without fully understanding how they function or assessing their reliability, and then often fail to test their reliability in use. Deficiencies in current disclosure laws make it impossible for the public to know if government algorithms are functioning properly or to identifying sources of ineffectiveness or bias.
- Algorithms can make government less accountable. Without informed oversight of the use of algorithms, humans can at once offload responsibility for their failures onto the algorithm, while also gaining unwarranted credibility due to the algorithm's perceived power of analysis.

These concerns are very real in Connecticut today, as confirmed by our efforts over the past year to assess the reliability and accountability of algorithms used by three State agencies, the Department of Children and Families (DCF), Department of Education (DOE), and Department of Administrative Services (DAS). Responses to Freedom of Information (FOI) requests confirmed both that existing disclosure requirements are insufficient to allow meaningful public oversight of the use of algorithms, and that agencies do not adequately assess the effectiveness and reliability of algorithms, either at the time of acquisition or after implementation. The FOIA responses generally revealed that agencies are insufficiently aware of the potential problems posed by their algorithms and unconcerned about the lack of transparency.

- DCF provided the only complete FOIA response, producing documents on its use of an algorithm intended to reduce the incidence of children suffering a life-threatening episode. It disclosed basic information about the algorithm but not its source code, which DCF did not possess and claimed to be protected as a trade secret. The production indicated that DCF had not performed a robust evaluation of the algorithm's efficacy or bias before implementing it or during the three years it was used.
- DOE made a partial production concerning its use of an algorithm to assign students to schools, an issue that has raised substantial disparate racial impact questions in the past. DOE's disclosure did not reveal how its school-assignment algorithm worked, apart from noting that it implemented the "Gale-Shapley deferred acceptance algorithm" and had no mechanism allowing parents to challenge its determinations. DOE refused to produce source code documenting operation of its algorithm on trade secret grounds, and provided no records related to its acquisition of the algorithm, beyond the procurement announcement and the contract (disclosing that it cost over $650,000 to acquire). DOE's incomplete production revealed no effort to evaluate the algorithm's efficacy or bias.
- DAS provided no documents in response to our request for information concerning a new algorithm used in hiring state employees and contractors. The agency did claim in a phone call that many of the requested documents would be withheld – we believe erroneously – under various FOIA exemptions, but despite a half-dozen follow-ups over a period of several months, the agency failed to produce and documents at all.

The problems with algorithmic accountability confirmed by this FOIA experiment call out for a legislative response. Several options have been implemented or seriously considered elsewhere, such as requiring: agencies to assess the effectiveness and bias of an algorithm, affirmative public disclosures about algorithms used to conduct government business, waiver of trade secret protection in certain circumstances, and mandatory disclosures to individuals subject to algorithmic decisions.

While the details of such approaches require study, it is imperative that steps are undertaken now to identify an effective response to the current lack of algorithmic accountability. The potential for serious harm to be inflicted by malfunctioning or biased algorithm is too serious to ignore.

INTRODUCTION

An algorithm is nothing more than a set of instructions for solving a problem or accomplishing a task, but algorithms lie at the heart of all computing. They are the DNA of the technological revolution that has changed the world and necessary for "machine learning," a form of artificial intelligence that allows computers to look through large volumes of data to identify patterns and develop novel ways to complete an assigned task. Everything from medical breakthroughs to climate change research requires algorithms and machine learning, as do the driving directions provided by Google maps and movie recommendations from Netflix. Increasingly, so too does government decision-making.

Governmental authorities at all levels are increasingly using algorithms and machine learning technologies that promise to make government function more effectively, efficiently, and timely. Governments already use algorithms for everything from screening job applications to assigning students to magnet schools, setting bail, and allocating social welfare benefits of all types.

The widespread adoption of algorithmic decision-making holds great promise, but also holds potential great peril. Without proper testing and ongoing evaluation, algorithms can function improperly or perpetuate historic biases reflected in the algorithm's code or, for a machine-learning algorithm, embedded in the data used to train it. Mistakes in making governmental decisions can inflict grave harm. Individuals' lives can be upended by an erroneous facial recognition match, faulty background check, incorrect denial of benefits, or wrongful termination of parental rights.

Yet, the public today often knows little about the use of algorithms in government decision-making, and government agencies often know just as little about whether the algorithms they use perform properly and without unintended consequences. Algorithms can seem like all-knowing black boxes that will correctly make the tough decisions and solve problems. Algorithms provide answers, but often without disclosing how the answers were reached. The public then is unable to assess whether an algorithm is working properly and whether its decision is fair and unbiased.

Our current transparency and accountability laws have not kept pace with the technological revolution and are not up to the task of providing the information needed for the public to understand the algorithms used by governments and to hold government agencies accountable for their use. Freedom of information laws have serious shortcomings when it comes to algorithmic accountability, largely because

companies that develop algorithms typically claim them to be trade secrets that cannot be disclosed to the public. They also often fail to generate the data necessary to evaluate how well an algorithm is working or whether it is biased.

This era of the algorithm is going to require new approaches to government transparency and accountability. Procedures need to be developed that will enable the public to know how the algorithms used to make government decisions are designed to work, whether they are performing as intended, and whether their decisions are unbiased and not simply automating past inequalities. Meaningful algorithmic accountability will require more than simply affording the public an opportunity to "look under the hood" of a software program. It will likely require some mandated process of expert assessment of government-used algorithms, accompanied by proper disclosure of those assessments to the public.

This paper is intended to raise awareness of the significant issues raised by our accelerating move toward government by algorithm,[1] and to begin a discussion about how best to address them.[2] The paper proceeds in three parts.

Part one documents a few case studies of government algorithms gone awry—software that has unknowingly perpetuated systemic bias, or that has made mistaken decisions with devastating impact on the lives of citizens. These examples show the gravity of the harms that can be inflicted by malfunctioning algorithms and highlight some important issues that need to be addressed when governments adopt algorithms: How does the public know what an algorithm is intended to do and measure if it is working properly? How does the public know if an algorithm's decisions are biased? Are human discretion and intervention incorporated into the algorithmic decision-making? And, is an algorithm cost-effective and procured through appropriate channels?

Part two presents the results of a yearlong effort to use the Freedom of Information Act ("FOIA") in Connecticut to obtain answers to those questions with respect to algorithms used by three different state agencies. This small-scale study

---

[1] *See, e.g.*, Cary Coglianese, *A Framework for Governmental Use of Machine Learning*, REPORT TO THE ADMIN. CONF. OF THE U.S., 3 (Dec. 8, 2020), https://www.acus.gov/sites/default/files/documents/Coglianese%20ACUS%20Final%20Report.pdf (noting that machine learning algorithms "have recently found themselves in use in so many products and settings that they appear to portend the reshaping of many important aspects of human life").

[2] Accordingly, this paper does not address private-sector algorithm use, which presents a different set of concerns.

illustrates some fundamental shortcomings of current transparency and accountability laws. Information about the algorithms took months to obtain if any information was provided at all. This experiment also revealed that some agencies knew little about the functioning of the algorithms they used, did little to confirm whether it performed as described by its creator, and did less to assess any possible bias in its outputs.

Part three provides a brief overview of some proposed legislative approaches to address the current lack of algorithmic transparency and accountability. Under various approaches, agencies would be required to assess the likely and actual effectiveness and bias of an algorithm, make certain affirmative disclosures to the public about their use of algorithms, require companies to waive trade secret protection for certain types of algorithms, and keep individuals informed about when they will be or have been subject to algorithmic decisions. Though the details of these approaches require further study, implementing any subset would contribute significantly to improving algorithmic transparency and accountability.

PART 1: THE GROWING PROBLEM OF ALGORITHMIC ACCOUNTABILITY

As technological capabilities have advanced, governmental agencies have increasingly delegated important administrative decisions to algorithms. Decisions concerning everything from hiring, to allocation of public benefits, and even law enforcement, often involve the use of algorithms today. These algorithms, however, frequently operate as a black box, making decisions without transparency and making it difficult or impossible to assess the accuracy and fairness of their performance. In contrast to the robust procedures that exist in many administrative contexts to promote the accountability of human decisions in government—from freedom of information laws to adjudicatory due process protections and appeals procedures—algorithms largely operate today in an accountability-free zone.

This lack of transparency and accountability has real and significant consequences: wasteful spending on inadequately tested automated systems, attenuated oversight of government operations producing biased or unfair results, and humans falling back on algorithms to avoid responsibility. For affected individuals, the consequences can literally be a matter of life and death. The data needed by private corporations for machine learning algorithms also raise significant, often unaddressed privacy concerns. More broadly, government reliance on algorithms that are poorly explained and little scrutinized may diminish public confidence in government.

The following discussion surveys some of the accountability deficits exposed by the spreading adoption of algorithms. It considers four areas in which significant

concerns are raised by the lack of accountability and transparency surrounding algorithms now in use: (a) algorithmic accuracy and effectiveness, (b) algorithmic bias, (c) algorithmic secrecy and procurement issues, and (d) human administration of algorithms (in particular, the extent to which humans administering algorithms exercise discretion).

A. Algorithmic Accuracy and Effectiveness

Algorithms, especially untested ones, can simply fail to do what they are supposed to do, with potentially catastrophic consequences for those subject to their decisions, not to mention the waste of taxpayer dollars. The real possibility of failure calls for rigorous oversight, as case studies involving criminal justice, child protective services, and welfare illustrate.

1. Criminal Justice and Policing

Without accountability and care surrounding the use of algorithms, agencies and municipalities can burn through millions of dollars before realizing a new technology simply does not work. For example, *Chicago*'s predictive policing program was designed to identify specific individuals who were more likely to be perpetrators of shootings, and anyone identified by the algorithm became subject to increased contact with the police. But the $2 million algorithm had no identifiable positive impact.[3] In *Palo Alto*, predictive policing software failed to reduce or solve crime and the software contract was cancelled.[4] In *Mountain View*, the police department discontinued a similar program when it concluded after five years of use that the results were "mixed."[5] *Los Angeles* also stopped using predictive policing algorithms when the department's inspector general was unable after a decade of use to determine that the software reduced crime.[6] In each of these cases, the resources could have been better spent elsewhere.

---

[3] Jessica Saunders et al., *Predictions Put into Practice: A Quasi-experimental Evaluation of Chicago's Predictive Policing Pilot*, 12 JOURNAL OF EXPERIMENTAL CRIMINOLOGY 347, 363 (2016); Matt Stroud, *Chicago's Predictive Policing Tool just Failed a Major Test*, VERGE (Aug. 19, 2016), https://www.theverge.com/2016/8/19/12552384/chicago-heat-list-tool-failed-rand-test.

[4] Mark Puente, LAPD Pioneered Predicting Crime with Data, Many Police Don't Think it Works, L.A. TIMES (July 3, 2019), https://www.latimes.com/local/lanow/la-me-lapd-precision-policing-data-20190703-story.html.

[5] *Id.*

[6] *Id.*

Some technology fails so catastrophically that it is shocking it was ever adopted in the first place. Facial recognition technology, for example, is notoriously ineffective in general use. Nevertheless, it is used more and more in policing, without adequate testing or training. In *Detroit*, for example, the police department faced public scrutiny after it wrongfully arrested several Black residents based on faulty facial recognition software.[7] The police department later admitted that its software misidentified suspects about "96 percent of the time."[8] Proponents of facial recognition software, including the software used in Detroit, argue that it should not be used alone, but as one of many tools for solving crimes.[9] But when the technology is misused without significant oversight – as it was in Detroit – it can ruin people's lives, especially those of Black residents, for whom the technology is least accurate.[10] *Florida* used another algorithm, a "risk score" algorithm that calculated a defendant's likelihood for recidivism and violence, that was found to be both incredibly ineffective and racially biased. Only 20 percent of those predicted to commit violent crimes went on to do so.[11] Time and again, criminal justice algorithms have been found to be ineffective and biased.

Using or experimenting with untested policing algorithms can also have negative consequences, posing a threat to public safety and to individuals. Experts note that ineffective policing software can "backfire" with unintended adverse effects. When individuals are wrongly targeted by the police, for example, they are less likely to trust the police in the future.[12] With a lack of trust, individuals are less likely to report crime and to use other crime-prevention resources, such as taking matters into

---

[7] Miriam Marini, Farmington Hills Man Sues Detroit Police After Facial Recognition Wrongly Identifies Him, Detroit Free Press (Apr. 13, 2021), https://www.freep.com/story/news/local/michigan/2021/04/13/detroit-police-wrongful-arrest-faulty-facial-recognition/7207135002.

[8] Timothy Lee, *Detroit Police Chief Cops to 96-percent Facial Recognition Error Rate*, ARS TECHNICA (June 30, 2020), https://arstechnica.com/tech-policy/2020/06/detroit-police-chief-admits-facial-recognition-is-wrong-96-of-the-time.

[9] *Id.*

[10] *See* Patrick Grother et al, *Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects*, INFORMATION ACCESS DIVISION, INFORMATION TECHNOLOGY LABORATORY, U.S DEPARTMENT OF COMMERCE, 7 (2019).

[11] Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[12] Brandon Welsh and Michael Rocque, *When Crime Prevention Harms: A Review of Systematic Reviews*, 10 J. EXPERIMENTAL CRIMINOLOGY 245 (2014); Joan McCord, *Cures That Harm: Unanticipated Outcomes of Crime Prevention Programs*, 587 Annals Am. Acad. Pol. & Soc. Sci. 16 (May 2003).

their own hands. Failed algorithms not only fail to reduce crime, they can make public safety tangibly worse.[13]

2. Child Protective Services and Foster Care

A child welfare algorithm produced by Mindshare Technology that used predictive analytics to estimate the chance that a child faced neglect, danger, or death has been used, for at least some time, by Connecticut, Illinois, and six other states.[14] Illinois dropped the algorithm after finding that it simultaneously overestimated the number of children at high chance of death and failed to identify several high-profile deaths—including one where the algorithm failed to flag a child who had been the subject of at least 10 child welfare abuse investigations.[15] The algorithm had other anomalous results, such as not linking the investigations of siblings that were part of the same household. Following a thorough review, Illinois dropped the software because "it didn't seem to be predicting much."[16] Resources at these agencies are thin, and both false positives and false negatives can create poor outcomes for children.

3. Welfare Programs and the War on Poverty

In 2016, Michigan wrongly accused Brian Russell of unemployment insurance fraud and seized his $11,000 tax refund check.[17] Russell was cleared of the charge in 2018, but in the intervening years was unable to provide for his family and ultimately filed for bankruptcy. The decision to accuse Russell was not made by a person, but by the Michigan Integrated Data Automated System (MIDAS), an algorithm that made roughly 48,000 accusations of fraud against unemployment insurance recipients.[18] Despite the millions of dollars Michigan spent on the software, the state's auditor

---

[13] *Id.*

[14] David Jackson & Gary Marx, *Data Mining Program Designed to Predict Child Abuse Proves Unreliable, DCFS Says*, CHICAGO TRIBUNE (Dec. 6, 2017), https://www.chicagotribune.com/investigations/ct-dcfs-eckerd-met-20171206-story.html.

[15] Connecticut's use of the algorithm is discussed in Part 2, *infra.*

[16] *Id.*

[17] Paul Egan, *State of Michigan's Mistake led to Man Filing Bankruptcy*, DETROIT FREE PRESS (Dec. 22, 2019), https://www.freep.com/story/news/local/michigan/2019/12/22/government-artificial-intelligence-midas-computer-fraud-fiasco/4407901002.

[18] Michele Gilman, *AI Algorithms Intended to Root out Welfare Fraud Often end up Punishing the Poor Instead*, CONVERSATION (Feb. 14, 2020), https://theconversation.com/ai-algorithms-intended-to-root-out-welfare-fraud-often-end-up-punishing-the-poor-instead-131625.

general reportedly found that 93 percent of MIDAS's fraud determinations did not involve any actual fraud.[19]

MIDAS was just one in a series of suspect algorithms used to determine eligibility for welfare recipients. In another Michigan case, the state used a very simple algorithm to disqualify those with outstanding felony warrants from receiving food assistance: the algorithm disqualified anyone who appears on a statewide database of individuals with outstanding felony arrest warrants.[20] A class-action lawsuit ultimately established that the algorithm wrongly disqualified Michigan citizens who qualified for assistance.[21] The algorithm made mistakes that likely could easily have been caught by a human decisionmaker, such as confusing someone with a relative convicted of a felony. When this occurred, misidentified individuals needed to pursue a laborious administrative process to clear their names. Such algorithmic errors can have catastrophic and life-altering consequences.

When benefits are wrongly terminated, an efficient and speedy appeals system is vital, but states often have failed to provide reasonable appeals processes out of misplaced trust in an algorithm. In many states, for example, individuals had little recourse after the adoption of algorithms to determine Medicaid eligibility resulted in wrongful termination, steep benefit cuts, and other catastrophic impacts to the elderly, the disabled, and their families.[22] In Indiana, Omega Young—a woman with ovarian cancer—was denied treatment after an algorithm determined she no longer qualified for benefits.[23] She lost her medication, food stamps, and transportation to Medicaid

---

[19] Paul Egan, *State Taps Unemployment Insurance Fund to Balance Books*, DETROIT FREE PRESS (Dec. 15, 2016), https://www.freep.com/story/news/local/michigan/2016/12/15/state-taps-unemployment-insurance-fund-balance-books/95416100; *Cahoo v. SAS Analytics Inc.*, 912 F.3d 887 (6th Cir. 2019).

[20] Tresa Baldas, *Court: Michigan Stiffed Deserving People out of Food Aid*, DETROIT FREE PRESS (Aug. 26, 2016), https://www.freep.com/story/news/local/michigan/2016/08/26/court-michigan-food-stamps-crime/89425014/.

[21] *Id.*

[22] Colin Lecher, *What Happens when an Algorithm Cuts Your Health Care*, VERGE (Mar. 21, 2018), https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy; Jen Fifield, *What Happens When States Go Hunting for Welfare Fraud*, PEW (May 24, 2017), https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2017/05/24/what-happens-when-states-go-hunting-for-welfare-fraud; *K.W. v. Armstrong*, ACLU OF IDAHO (Sep. 15, 2016), https://www.acluidaho.org/en/cases/kw-v-armstrong.

[23] Emma Bowman, *'Automating Inequality': Algorithms in Public Services Often Fail the Most Vulnerable*, NPR (Feb. 19, 2018),

appointments. By the time Young won her administrative appeal for wrongful termination, she had been dead for a day.[24]

B. Algorithmic Bias

While algorithms may provide significant efficiency advantages over human decision-making, absent thorough assessment and outcomes testing, their use risks injecting bias into agency decisions. This risk was also made clear in case studies involving policing and child protective services algorithms.

1. Criminal Justice and Policing

Police departments around the country have adopted algorithmic tools to fight crime that frequently have had easily predictable racially disparate impacts.

For example, "hotspot" policing algorithms may lead to racially disparate encounters with the police. These algorithms are designed to predict the areas where there is a historically high rate of arrests, and are used by police departments to allocate their forces. However, a Black American is five times more likely to be stopped by police than a white American, and twice as likely to be arrested. As a result, an algorithm that uses arrest data as a baseline for police deployment decisions may predispose areas with higher shares of Black residents to more intense policing—and evidence suggests that some hotspot algorithms actually do have that effect.[25] While there is nothing intrinsically problematic about policing areas with higher crime rates, using these algorithms has the unintended effect of making the police in the field more suspicious in these areas. Specifically, algorithms may "prime" police to think they are in greater danger of encountering criminals in areas the algorithm identifies as a hotspot, which results in more invasive policing than in other communities—residents in these communities are subject to aggressive police encounters that a majority-white neighborhood would not face.[26]

---

https://www.npr.org/sections/alltechconsidered/2018/02/19/586387119/automating-inequality-algorithms-in-public-services-often-fail-the-most-vulnerab.

[24] *Id.*

[25] Will Douglas Heaven, *Predictive Policing Algorithms Are Racist, They Need to Be Dismantled*, MIT TECH. REV. (July 17, 2020), https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice.

[26] *Id.* (citing evidence suggesting that a hotspot algorithm "primes officers to expect trouble when on patrol, making them more likely to stop or arrest people there because of prejudice rather than need").

The policing algorithms thus create a racist "runaway feedback loop."[27] Research has shown that predictive policing algorithms that identify high-crime areas inflate the crime rates in the communities it identifies as hotspots.[28] This occurs both because police sent to a particular region will be more likely to see crime there, and because the absence of police in areas designated as lower crime will report less crime in those areas than actually occurs.[29] This skewed data gets fed back into the algorithm, which in turn predicts an ever higher relative crime rate in the area already identified as a hotspot. The problem iterates, resulting in a situation where, "even if crime rates in two neighborhoods are remarkably similar, if region A has a crime rate of 10% and region B has a crime rate of 11%, the update process will settle on region B with 100% probability."[30] Given that pre-algorithm baseline crime rates are likely inflated for Black communities, use of these algorithms will predictably amplify racially disparate policing.

This problem is exacerbated by the data some policing algorithms use to classify areas as high-risk. For example, many policing algorithms use arrests or 9-1-1 calls as an indicator of crime, rather than actual convictions. A report by the Chicago Office of the Inspector General found that the city's police department "overly relied on arrest records to identify risk even where there was no further arrest or arrests did not lead to convictions."[31] Predictably, the poorest and least white areas of the city had the highest number of emergency calls.

The bias in these algorithms was recently confirmed in a study conducted by two news organizations. They began by analyzing the impact of PredPol, discussed in Part 1.A.1, above, purchased by police in Plainfield, New Jersey in 2018 to predict where crime would likely happen. An analysis of its operation revealed stark bias

---

[27] Danielle Ensign et al, *Runaway Feedback Loops in Predictive Policing*, 81 PROCEEDINGS OF MACHINE LEARNING RESEARCH 1 (2018).

[28] *See* Kristian Lum and William Isaac, *To Predict and Serve?*, SIGNIFICANCE (2016), https://www.fatml.org/schedule/2016/presentation/predict-and-serve.

[29] *See* Renata M. O'Donnell, *Challenging Racist Predictive Policing Algorithms Under the Equal Protection Clause*, 94 N.Y.U. L. REV. 545, 563-64 (2019), https://www.nyulawreview.org/wp-content/uploads/2019/06/NYULawReview-94-3-ODonnell.pdf.

[30] *Id.* (internal quotation marks omitted).

[31] Tim Lau, Predictive Policing Explained, Brennan Center (Apr. 1, 2020), https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained; Joseph Ferguson, *Advisory Concerning the Chicago Police Department's Predictive Risk Models*, CITY OF CHICAGO OFFICE OF INSPECTOR GENERAL (Jan. 2020), https://igchicago.org/wp-content/uploads/2020/01/OIG-Advisory-Concerning-CPDs-Predictive-Risk-Models-.pdf.

resulting in a skewed allocation of resources caused by use of the algorithm. For example, comparing two neighborhoods of roughly equal size less than a mile apart over the same period of time, the software predicted 11 crimes in the more predominately white neighborhood and predicted 1,940 crimes in the neighborhood with more minorities. The news organizations then analyzed more than 5 million predictions made by the software, and it consistently predicted less crime in populations with a higher percentage of white residents.[32] In these cases, the underlying data used by the algorithm compounds existing inequalities, even when there are no criminal convictions that would provide the best evidence of whether an area is high-crime.

2. Child Protective Services and Foster Care

Studies have repeatedly shown that Black and Native American children are far more likely to have contact with Child Protective Services (CPS) than white children. Black and Native American families are almost twice as likely as white families to face a CPS investigation, foster care, and termination of parental rights.[33] In some states, Native American children are four to eleven times more likely to be placed in foster care than white children.[34] In an attempt to combat these disparities, some reformers have argued that using algorithms will reduce the racial disparities present in the child welfare system.[35] In one high-profile case, however, biased algorithms exacerbated the problem.

---

[32] Aran Sankin, et. al., *Crime Predictions Software Promised to be Free of Biases. New Data Shows It Perpetuates Them*, MARKUP (Dec. 2, 2021), https://themarkup.org/prediction-bias/2021/12/02/crime-prediction-software-promised-to-be-free-of-biases-new-data-shows-it-perpetuates-them.

[33] Emily Putnam-Hornstein et al., *Cumulative Rates of Child Protection Involvement and Terminations of Parental Rights in a California Birth Cohort, 1999–2017*, 111 AM. J. PUB. HEALTH 1157 (2021), https://ajph.aphapublications.org/doi/pdf/10.2105/AJPH.2021.306214; *New Study Confirms High Prevalence of Investigations, Loss of Parental Rights,* IMPRINT (APR. 22, 2021), https://imprintnews.org/youth-services-insider/study-high-prevalence-investigations-loss-parental-rights/53687; Emnet Almedom et al., *Algorithms and Child Welfare: The Disparate Impact of Family Surveillance in Risk Assessment Technologies*, BERKELEY PUB. POL'Y J. (Feb. 2, 2021), https://bppj.berkeley.edu/2021/02/02/algorithms-and-child-welfare-the-disparate-impact-of-family-surveillance-in-risk-assessment-technologies.

[34] *Disproportionate Representation of Native Americans in Foster Care Across United States*, POTAWATOMI NATION (Apr. 6, 2021), https://www.potawatomi.org/blog/2021/04/06/disproportionate-representation-of-native-americans-in-foster-care-across-united-states; Stephen Pevar, I*n South Dakota, Officials Defied a Federal Judge and Took Indian Kids Away From Their Parents in Rigged Proceedings*, ACLU (Feb. 22, 2017), https://www.aclu.org/blog/racial-justice/american-indian-rights/south-dakota-officials-defied-federal-judge-and-took.

[35] Daniel Heimpel, *Can Predictive Analytics Root Out the Social Workers Most Likely to Break up Black Families?*, IMPRINT (June 19, 2019), https://imprintnews.org/news-2/can-predictive-analytics-root-

The Allegheny Family Screening Tool (AFST), a new tool used in Pennsylvania, has faced criticism for bias and effectiveness.[36] For one, the tool has a consistent and clear bias against the poor. In part, this is due to the data on which the algorithm's decision-making is based. Much of the data comes from a family's interaction with public agencies. The AFST uses an average of 800 records for each resident of the county; however, some residents have far more data than others. For example, the algorithm uses poverty-related variables when calculating a child's risk factor for neglect. However, the factors to determine neglect – including lack of healthcare, food, homelessness, or a parent's absence at work – frequently overlap with the consequences of poverty.[37] Indicators of neglect and poverty can overlap, but there is a large difference between deliberate abuse and abject poverty. One report concluded that "AFST mostly just reports how many public resources families have consumed."[38] Medicaid, food stamps, SSI, and geographic neighborhood data (correlated with poverty) are among the many poverty-related variables used by the AFST to calculate a child's risk score.[39] After all, the government has the most data on those types of families most likely to come in contact with government agencies.[40] The algorithm's reliance on this data thus created a substantial risk that it discriminated against the poor.[41]

---

out-removal-happy-social-workers/35650; Susan Wandalowski, *Your View by Northampton Human Services Department: How an Algorithm Will Help Keep Families Together*, MORNING CALL (Apr. 11, 2021), https://www.mcall.com/opinion/mc-opi-software-tool-child-welfare-wandalowski-20210411-h4qw5whgura2faetmcohrrkjva-story.html.

[36] Virginia Eubanks, *A Child Abuse Prediction Model Fails Poor Families*, WIRED (Jan. 15, 2018), https://www.wired.com/story/excerpt-from-automating-inequality.

[37] *Id.*

[38] *Id.*

[39] *Id.*

[40] *Id. See also* Kim Strong, *Computer Program Aimed at Predicting Child Abuse Debated*, YORK DAILY REC. (July 10, 2021), https://apnews.com/article/child-abuse-b9e948bba66efe03b9a06a023f2e00b8 ("The data that's used for algorithms is primarily public data, so the poor family using government services for food, housing, drug and alcohol counseling and mental health treatment will have much more data in the public sphere than a wealthier family using private insurance for counseling and treatment.").

[41] It bears noting that these criticisms of the AFST were possible only because the algorithm was more transparent than most algorithms provided by private-sector companies. *See* Part 1.C, *infra*.

C. Lack of Transparency in the Acquisition and Operation of Algorithms

A general lack of transparency surrounding the decision to acquire an algorithm compounds already significant concerns about the effectiveness and bias of algorithms used in government. Both the procurement procedures for algorithm acquisition and the contents of algorithms themselves are often shrouded in secrecy. Procurement problems arise when states and localities have acquired algorithms in ways that bypass the normal processes. Transparency in procurement promotes competition among vendors, requires government justification of its purchasing decisions, and invites public scrutiny, making the adoption of ineffective or dangerous algorithms less likely. Procurement secrecy, by contrast, facilitates the deployment of algorithms without adequate vetting and justification. Likewise, trade secret protections prevents the public, and sometimes even the agency, from evaluating if the algorithm is biased or ineffective.

Illinois's decision in 2017 to abandon the Mindshare algorithm, discussed above, provides a window into the importance of procurement procedures. Illinois had adopted the proprietary algorithm under questionable circumstances—a no-bid contract that the Illinois Department of Children and Family Services (DCFS) misclassified as a grant, so that it "avoided state bidding transparency requirements, making it impossible to determine if Illinois could have obtained the same services from local companies at a lower cost, a requirement of the state's procurement."[42] It abandoned the untested algorithm a year later.[43]

New Orleans's experience with a predictive policing algorithm provides another example of how bypassing typical procurement processes can inhibit algorithmic accountability. Palantir, a firm that has attracted controversy over its aggressive efforts to amass data on U.S. citizens,[44] offered New Orleans its experimental predictive policing software for free in order to give Palantir an opportunity to test the product on New Orleans citizens. The New Orleans police department and the mayor's office entered into an agreement with Palantir to use its algorithm without public disclosure or consultation with other stakeholders. This

---

[42] David Jackson & Gary Marx, *Data mining program designed to predict child abuse proves unreliable, DCFS says*, CHICAGO TRIBUNE (Dec. 6, 2017), https://www.chicagotribune.com/investigations/ct-dcfs-eckerd-met-20171206-story.html.

[43] *Id.*

[44] Ali Winston, *Palantir Has Secretly Been Using New Orleans to Test Its Predictive Policing Technology*, THE VERGE (Feb. 27, 2018), https://www.theverge.com/2018/2/27/17054740/palantir-predictive-policing-tool-new-orleans-nopd.

relationship apparently led New Orleans to give Palantir free access to criminal and non-criminal data to train its software, a move that could undermine its citizens' privacy interests.[45] By characterizing the acquisition of the algorithm as philanthropic rather than commercial, they evaded the ordinary procurement process. Secrecy about the acquisition and deployment of the algorithm inhibited ordinary channels of accountability so much that, when *The Verge* contacted New Orleans city council members in 2018, after the program had been operating for six years, none of them knew of the program's existence.[46] The city then announced it was abandoning the algorithm a mere two weeks after *The Verge*'s story broke.[47] This example illustrates the importance of transparency about the acquisition and use of algorithms to facilitate scrutiny and monitoring, both by the public and by regulators.

These examples likely are not isolated. Individual agencies are the natural testing ground for algorithms designed for government use, as data is needed to test the algorithm. The first agencies to use an algorithm will justifiably be skeptical of their efficacy. These facts combined may well lead agencies to serve as pilots for the algorithm at no cost, thereby bypassing ordinary procurement processes—as discussed in Part 2, this appears to be how Connecticut acquired use of the Mindshare algorithm. While this may be a financial boon for an agency, lack of oversight prevents the public from evaluating any flaws the untested algorithm might have.

Aside from such procurement irregularities, the commercial acquisition of algorithms by government agencies can pose fundamental obstacles to transparency due to their proprietary nature. Secrecy about the workings of an algorithm usually flows from trade secret claims made by developers of proprietary algorithms and makes monitoring the performance of algorithms particularly difficult. If a vendor claims trade secret in an algorithm, its workings will be shielded from disclosure under freedom of information laws.[48] A lack of information about the factors used by an algorithm impedes efforts to assess their reliability and effectiveness, and thus some

---

[45] *Id.*

[46] *Id.* (noting that "[t]he Palantir partnership would have likely received more scrutiny from the city council had it been itemized in a budget").

[47] Ali Winston, *New Orleans Ends its Palantir Predictive Policing Program* THE VERGE (Mar. 15, 2018), https://www.theverge.com/2018/3/15/17126174/new-orleans-palantir-predictive-policing-program-end.

[48] *See* Part 2, *infra* (discussing proprietary algorithms in use by Connecticut agencies).

scholars argue that algorithms used by state agencies, particularly in the criminal context, should necessarily be subject to public disclosure.[49]

Trade secret protection may be so strong as to prevent the government from performing necessary oversight. For instance, the Dutch Ministry of Social Affairs and Employment used an algorithm called System Risk Identification (SRI) to help identify meritless claims for unemployment benefits, but the algorithm's proprietary nature prevented meaningful investigation of its efficacy: "In 2019, a city council hearing with the social ministry abruptly ended when members of the city government wouldn't sign nondisclosure agreements before receiving a briefing about how the system works."[50] Without access to even general information about the criteria on which an algorithm relies and the underlying data on which it draws, human oversight is unlikely to be effective.

A lack of transparency about the criteria on which on which an algorithm relies certainly impedes efforts to promote algorithmic accountability. The Allegheny Family Screening Tool, discussed above,[51] is more accountable than other child protection algorithms in large part because of the transparency of its code: "It is owned by the county. Its workings are public. Its criteria are described in academic publications and picked apart by local officials."[52] While transparency alone is not sufficient to yield adequate algorithmic accountability, its absence creates a substantial risk that an algorithm will be inadequately tested and evaluated. It is important to caution, however, that operational biases are not likely to be visible from an algorithm's source code alone, and broader scrutiny of its operation will be required to detect disparate impacts.[53] Indeed, the AFST still likely produces disparate impacts due to biases in the

---

[49] *See* Alyssa M. Carlson, *The Need for Transparency in the Age of Predictive Sentencing Algorithms*, 103 IOWA L. REV. 303 (2017), https://ilr.law.uiowa.edu/print/volume-103-issue-1/the-need-for-transparency-in-the-age-of-predictive-sentencing-algorithms/.

[50] Cade Menz & Adam Satariano, *An Algorithm That Grants Freedom, or Takes It Away*, N.Y. TIMES (Feb. 6, 2020), https://www.nytimes.com/2020/02/06/technology/predictive-algorithms-crime.html.

[51] *See* Part 1.B.2, *supra*.

[52] Dan Hurley, *Can an Algorithm Tell When Kids Are in Danger?*, N. Y. TIMES (Jan. 2, 2018), https://www.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html.

[53] *See* Part 1.D, *infra*.

data it draws on.[54] Nevertheless, transparency about the algorithm's workings makes it easier to corroborate such a claim.

The efficacy of human oversight will largely depend on the extent to which overseers can deploy techniques to assess the accuracy and fairness of the operation of algorithms.[55] Doing so depends on at least minimal transparency about algorithmic criteria, the data on which the algorithm draws, and the output of the algorithm. Secrecy about these factors will undermine the ability of human monitors to pursue accountability in meaningful ways. Meaningful accountability will not necessarily require absolute transparency about algorithms' source code, as long as sufficient information is provided to monitors to enable meaningful evaluation.[56] For instance, it might be sufficient for the state to provide a description of the factors that the algorithm considers rather than the full source code. But trade secret protection can mean that algorithmic criteria are described in such vague terms that it is difficult to draw meaningful inferences about an algorithm's workings.[57] And a description of the factors on which an algorithm relies will not be sufficient to detect biases arising from training data.

D. The Human Element: Administrative Discretion

Algorithmic secrecy is a significant obstacle to the accountable and fair administration of algorithms. Absent transparency, human decision-makers lack information they need to scrutinize the performance of algorithms, assess the reliability of their output, and justify reliance upon them. But there can also be accountability deficits arising from procedural ambiguity and non-transparency about the way agencies deploy algorithms. These deficits arise largely from the public perception of algorithms as accurate decision-makers, allowing humans to offload responsibility for their own discretionary decisions.

Predictive policing, an area in which many municipalities have made use of algorithms, provides illustrative examples of the defects introduced by human discretion. Human administration of an algorithm can introduce additional biases and caprice into its execution, and conversely the halo provided by an algorithm can both

---

[54] *See* Part 1.B.2, *supra*.

[55] Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PENN. L. REV. 633, 637 (2017), https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=9570&context=penn_law_review.

[56] *Id. See also* Part 3, *infra* (discussing proposals for independent oversight of government algorithms).

[57] *See* Part 2.B, *infra* (discussing the limitations of the information Connecticut's Department of Education provided regarding the factors considered by its Regional School Choice System).

disguise arbitrary administrative action and encourage humans to rely on the algorithm more than is justified.[58] Illustrating all these problems, Chicago's Strategic Subject List, known colloquially as the "heat list," purported to identify individuals at heightened risk of involvement in shootings. Police invoked the algorithm to gain credibility, even though it turned out not to be particularly accurate and to do little to constrain human discretion.[59] In effect, "[c]ops would just use the list as a way to target people."[60] A review by the Inspector General of the Los Angeles Police Department found significant disparities in the administration of the LAPD's predictive policing algorithms, PredPol and LASER, by different departments across the city.[61] As a result of the investigation, LAPD stopped using LASER in 2019.

Similar problems occur in other contexts where procedures permit human discretion to supplant algorithmic recommendations. For instance, risk assessment tools implemented to guide bail determinations are only as effective as the judgment of the human beings who administer them.[62] If human administrators have the power simply to overrule an algorithm's recommendation, then administrative decisions may fail to reflect the performance of the algorithm, instead reflecting the very flaws in judgment that algorithmic decision-making is supposed to mitigate. Even if the algorithm provides highly accurate recommendations, the results of the process in which it is used may be defective simply because people can disregard it. Moreover, when people overrule algorithmic recommendations at a high rate, it is difficult to judge the performance of the algorithm or to assess its value as an input into human

---

[58] John Zerilli et al., *Algorithmic Decision-Making and the Control Problem*, 29 MINDS AND MACHINES 555 (2019); Angelika Adensamer et al., *"Computer Says No": Algorithmic Decision Support and Organisational Responsibility*, J. RESPONSIBILITY TECH. (2021), https://www.sciencedirect.com/science/article/pii/S266665962100007X; DAVID FREEMAN ENGSTROM ET AL., GOVERNMENT BY ALGORITHM: ARTIFICIAL INTELLIGENCE IN FEDERAL ADMINISTRATIVE AGENCIES (2020), https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf.

[59] Jessica Saunders, Priscilla Hunt & John S. Hollywood, *Predictions Put into Practice: A Quasi-Experimental Evaluation of Chicago's Predictive Policing Pilot*, 12 J. EXPER. CRIM. 347 (2016).

[60] Matt Stroud, *Heat Listed*, THE VERGE (May 24, 2021), https://www.theverge.com/22444020/chicago-pd-predictive-policing-heat-list.

[61] Martin Macias Jr, *Audit Finds LAPD Predictive Policing Programs Lack Oversight*, COURTHOUSE NEWS SERVICE (Mar. 8, 2019), https://www.courthousenews.com/audit-finds-lapd-predictive-policing-programs-lack-oversight/.

[62] *See* Megan Stevenson, *Assessing Risk Assessment in Action*, 103 MINN. L. REV. 303, 373 (2018), https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1057&context=mlr (observing that "judges . . . deviated from the action-directives associated with the risk assessment more often than not").

judgment.[63] In such contexts, the use of an algorithm may provide unwarranted legitimacy, if the algorithm is thought to improve accuracy and reduce bias but in fact the algorithm fails to constrain human discretion in practice.

PART 2: OBSTACLES TO ALGORITHMIC ACCOUNTABILITY UNDER CURRENT FOI LAW: A CONNECTICUT STUDY

Connecticut is not immune to possible algorithmic pathologies, such as potential bias and ineffectiveness, lack of transparency, and problematic discretion. To assess the extent to which current transparency laws are sufficient to expose such problems, we submitted Freedom of Information (FOI) requests for certain information about algorithms known to be used by Connecticut's Department of Children and Families (DCF), Department of Education (DOE), and Department of Administrative Services (DAS). Through these requests we sought to learn basic information about (1) what the algorithms do, (2) how the algorithms were acquired, (3) how the agency tested or assessed the algorithm, and (4) the extent to which answers to these questions can be obtained under current disclosure laws. The requests and the agencies' complete responses are available at the following website: https://law.yale.edu/mfia/projects/government-accountability/algorithmic-accountability.

As discussed below, the agencies' responses to our FOIA requests were generally deficient across all four metrics. In one case this was due to a disturbing dearth of information possessed by the agency; in others, our analysis of the first three factors was hampered by the agency's failure to provide a complete—or, in one case, any—response to our FOIA requests despite persistent follow-ups. We present the following analysis of the responses not to suggest wrongdoing or to discredit the work of these agencies, but to illustrate the insufficient assessment and understanding of algorithms used to perform governmental functions that is currently endemic at all levels of government. This may well be a symptom of the current legal environment, some possible cures for which are discussed in Part 3.

A. Department of Children and Families (DCF)

On April 5, 2021, we requested records from DCF concerning the use of algorithms in the identification of at-risk children.[64] We were aware that DCF had used the Eckerd Rapid Safety Feedback (ERSF) model for child welfare interventions, which incorporates the Mindshare Technology child risk assessment algorithm

[63] *Id.* at 369.

[64] DCF FOI Request.

discussed in Part 1. We were also aware that Illinois had abandoned both ERSF and the Mindshare algorithm due to their ineffectiveness.[65] We sought to learn the circumstances around the algorithm's acquisition and termination by DCF, how the algorithm worked (including the extent of human involvement), what factors it used to trigger a recommendation for child welfare action, and whether DCF assessed the algorithm for bias and efficacy before or during its use. DCF issued an initial response five months after receiving the request. Several requests were unanswered and a follow-up inquiry was promptly submitted, which yielded a supplemental response two months later.

DCF produced fewer than 200 pages of responsive documents. The material consisted largely of slideshows and a manual created by Eckerd that was apparently used for training, and perhaps also for marketing. DCF also produced its contract with Eckerd, a termination letter, some brief, general evaluations of ERSF (mostly conducted by Eckerd), and additional documents providing some further information about ERSF. These documents contain some basic information about the ERSF model and how the Mindshare algorithm fits into it.

The records show, for example, that DCF obtained the software to "reduce the incidence of children (or siblings) known to the Department who experience an intake on a new or reopened case not already transferred to Ongoing Services, followed at any time by a subsequent substantiated life-threatening episode."[66] The records also reveal a bit about how the software was used. According to a flowchart generated by Eckerd, a DCF Careline worker first screens calls for possible intervention.[67] Data from screened-in cases were then plugged into the Mindshare algorithm,[68] which would examine data associated with the child to calculate a "similarity" score for that child relative to other children who had previously experienced poor outcomes.[69] If a similarity score exceeded 50 percent, Eckerd and DCF would develop a plan for the child's care.[70] If the similarity score was below 50 percent DCF would process the

[65] The ERSF model also includes non-algorithmic components such as quality assurance, coaching, and mentoring for DCF staff. *See* DCF Production at 6.

[66] *Id.* at 7.

[67] *Id.* at 6.

[68] *Id.* at 6.

[69] *Id.* at 59-60.

[70] *Id.* at 6, 59-60.

case without using ERSF to develop a follow-up plan.[71] We were therefore able to assess from the publicly available records how the agency used the Mindshare algorithm.[72]

The responsive records, however, provide little to no information about several issues, such as testing of the algorithm and assessments of any bias, suggesting that DCF did little to evaluate the working of the Mindshare algorithm before or during its use. The available records also contain little information about the internal logic used by the algorithm to determine its "similarity score." DCF did not produce the algorithm's source code that we specifically requested because it "did not host the software and does not have access to the source code or algorithm, "[73] and furthermore the source code was "proprietary to Mindshare" and "not disclosable under the Freedom of Information Act."[74]

DCF's production did provide some insight into the data used in generating the similarity score, noting that the algorithm regularly updates the specific factors it uses to consider a child's risk using supervised machine learning.[75] DCF produced a screenshot of the twenty-five most important factors used at one point in time,[76] but because the algorithm adjusts factors on an ongoing basis there is no way to know if these factors are representative of the factors used during the three years DCF used the algorithm. DCF's screenshot also contains a rank order of the weights assigned to each of the factors then in use but does not disclose the actual weights assigned to each factor by the algorithm.

DCF's records provided little information about *how* the algorithm's similarity score is updated over time, including which children's data are used to create the similarity metric. It is therefore unclear if the algorithm's training data is representative of the population, which would be necessary to stave off bias and prevent the algorithm from generating a significant number of false negatives or false

---

[71] *Id.* at 6.

[72] The precise dates that DCF employed the algorithm are unclear. DCF's cover letter to our request states that the algorithm was in use from approximately October 2016 through August 2019. *Id.* at 1. However, the contract between DCF and Eckerd took effect on July 18, 2016, and a letter from DCF to Eckerd purports to terminate their relationship on December 31, 2019. *Id.* at 102, 188.

[73] *Id.* at 187.

[74] *Id.*.

[75] *Id.* at 58.

[76] *Id.*

positives. Both are significant in child protective services. False negatives mean that an abused or neglected child will not get needed help, while false positives result in unwarranted government intrusion into children's relationships with their parents.

It is also unclear whether the algorithm caused these or any other negative effects because, in DCF's words, it was "not able to complete a robust formal evaluation to assess [ERSF's] efficacy."[77] DCF did produce an email from one of its employees providing a "preliminary" evaluation that "does not support the effectiveness of [ERSF] on maltreatment recurrence."[78] This preliminary evaluation apparently conflicted with a prior evaluation done by Eckerd that did not purport to assess the algorithm specifically and which DCF did not produce beyond references to its existence in the same email chain. DCF also produced a review by Eckerd of aspects of the operation of ERSF from March-June 2019 that did not address the impact of ERSF or the Mindshare algorithm on DCF's performance. DCF produced nothing to indicate that it, Eckerd, or Mindshare ever evaluated the algorithm for possible bias or disparate impact.

Finally, DCF provided little information about either its acquisition or termination of ERSF or the Mindshare algorithm. We requested procurement-related records, including any requests for proposals, and received none (apart from DCF's contract with Eckerd). This suggests that DCF may have bypassed ordinary procurement channels – and indeed, its contract with Eckerd provides that Eckerd and another outside entity, Casey Family Programs, covered the costs of DCF's use of the program from March 7, 2016 to September 30, 2018.[79]

DCF did produce a letter from its contracts management division to Eckerd terminating Eckerd's services effective December 31, 2019. The letter indicated that the termination was caused by DCF's resource constraints and did not reflect any dissatisfaction with the algorithm or Eckerd's services.[80]

In sum, FOIA allowed us to obtain some useful, basic information about DCF's use of the Mindshare algorithm, but that information is insufficient for meaningful public oversight. Little can be discerned through FOIA about how the algorithm works or its internal logic: DCF does not possess the source code—and would otherwise have refused to produce it due to its proprietary status—and did not

[77] *Id.* at 187.

[78] *Id.* at 189.

[79] *Id.* at 102.

[80] *Id.* at 188.

produce evidence of any careful evaluation of the effects of the algorithm. DCF used the algorithm for three years without performing a robust evaluation of its effectiveness, but relied instead on assessments by the vendor, which diverged from the one preliminary evaluation DCF did perform. Nobody evaluated whether the algorithm exhibited bias. And circumstances surrounding the algorithm's procurement remain a mystery.

B. Department of Education (DOE)

Whereas the DCF response illustrates the need for new transparency rules if there is to be effective oversight of public sector algorithms, the other two responses underscore inadequacies of FOIA generally.

We submitted a FOIA request to DOE on April 5, 2021, requesting information about DOE's procurement, use, and assessment of algorithms in the administration of school lotteries.[81] We selected DOE as an agency of interest because the use of algorithms in school assignment elsewhere has raised serious issues of disparate impact.[82] We were interested to learn whether and how DOE has taken precautions against disparate impacts and how it has assessed the performance of its algorithm in this respect. This was of particular importance given the recent settlement agreement in *Sheff v. O'Neill*, a case challenging disparate spending between schools in majority-white and majority-minority areas of Connecticut.

DOE confirmed receipt of our FOI request on April 12, warning "that it may take significant time to collect and review responsive records." We heard nothing further for several months.

We wrote to DOE on September 2 to inquire about the status of our request. DOE responded that "we anticipate that we should have a partial response available for you by the end of next week." After hearing nothing, we followed up two weeks later, on September 15, asking for a status update. DOE responded saying this time that the Regional School Choice Office "will try to provide a partial response within the next few weeks." When we still had received nothing a few weeks later, we followed up again on October 11. We mentioned that, in view of DOE's failure to produce records in a timely manner, we might need to pursue alternative avenues.

---

[81] DOE FOI Request.

[82] *See, e.g.*, Colin Lecher & Maddy Varner, *NYC's School Algorithms Cement Segregation. This Data Shows How.*, MARKUP (May 26, 2021), https://themarkup.org/news/2021/05/26/nycs-school-algorithms-cement-segregation-this-data-shows-how (presenting data showing that "Black and Latino students are regularly screened out of high schools across New York City—most strikingly, the city's top-performing schools").

DOE responded the next day saying "[w]e are hoping to have it completed by early next week."

On October 14, we did receive a "partial response" of non-exempt documents "concerning the Student Assignment Plan for voluntary inter-district schools within the Greater Hartford Region pursuant to the Phase IV Stipulation and Order from January 10, 2020 in *Sheff v. O'Neill*."[83] A cover message said that DOE "will provide additional responses as soon as possible."[84] As of the date of this report, two months later, we have received no further response.

The documents produced on October 14 included the procurement announcement, the purchase order (indicating that Connecticut paid $657,825 for the algorithm), the statement of work provided by Blenderbox (indicating that the algorithm was bespoke and including some promotional materials from Blenderbox), a Q&A addressing the algorithm's compliance with the *Sheff* agreement (with answers that are quite vague[85]), and the court order in *Sheff v. O'Neill* (requiring that Connecticut take steps to facilitate Hartford-resident minority students' attendance at magnet and technical schools). DOE also adverted to the online Family Guide provided to lottery participants (https://portal.ct.gov/-/media/SDE/School-Choice/RSCO/RSCOFamilyGuide.pdf) and indicated that there is no process for families to challenge placements through the Regional School Choice System.

The partial set of documents contained nothing that would indicate the extent to which DOE oversees the algorithm it uses. Nor, thus far, has DOE provided any information about the source code of the algorithm or the data it draws on. DOE informed us that it was withholding "[d]ocuments detailing coding of the placement system" as well as "[d]ocuments regarding specific data inputs and protocols," invoking FOIA's trade secret exemption.[86] DOE also invoked the trade secret exemption to withhold "[d]ocuments describing placement protocols."[87] DOE did

---

[83] DOE Production at 1.

[84] *Id.*

[85] For example, in response to a question about placement order, DOE answers that "[p]lacement will be . . . in a manner consistent with the stipulated agreement." *Id.* at 99. In response to a question about additional factors that the algorithm might incorporate, DOE answers, "SDE will review the data and as appropriate consult with its expert to determine if, in the future, additional variables will inform SES Tier assignments." *Id.* at 105. In response to a question about what training will be provided to staff, DOE responds, "[t]his is being developed." *Id.* at 106.

[86] *Id.* at 4-5.

[87] *Id.* at 4.

not elaborate about the extent of information it was withholding under these categories, but so far has provided almost no information about the functioning of the algorithm. The Blenderbox statement of work indicates only that the algorithm "is powered by an implementation of the Gale-Shapley deferred acceptance algorithm."[88]

The information that DOE provided about the factors considered by the algorithm is just the rudimentary information contained in the Family Guide. It reports that "[k]ey factors in assigning placements are the number of seats available, the grade level, and the applicants' choices, along with socioeconomic factors (SES), student preferences, priorities (such as sibling, staff and neighborhood), pathways and other factors."[89] This description provides little specificity about the meaning of these factors and no information at all about the weight they are given by the algorithm. While it is clear that some factors—in particular socioeconomic factors and neighborhood priorities—could give rise to disparate impact issues, DOE provided an insufficient explanation to know whether the algorithm's use of these factors does contribute to a disparate impact. And the description of the factors is itself incomplete, as the Student Guide fails to identify the "other factors" it says the algorithm may consider.

DOE also provided no information about any procedures it has used to assess the performance of the algorithm, including of the algorithm's effectiveness and any potential disparate impacts caused by its use, even though our FOI request specifically sought this information. DOE's failure to produce it may be because the agency has not conducted the assessments. Indeed, the Q&A regarding compliance with the *Sheff* agreement that DOE provided suggests that it has not yet devised the criteria on which the algorithm's performance would be evaluated.[90] It would obviously be concerning if DOE has implemented the algorithm with no monitoring of its performance, but it is impossible to know from DOE's FOIA response.

Our inability to understand and assess the functioning of the algorithm, due to DOE's broad invocation of the trade secret exemption and its insufficient response, illustrates the need for a new approach to providing algorithmic transparency.

---

[88] *Id.* at 15.

[89] *Id.* at 4. Regional School Choice Office, Family Guide to School Choice in the Greater Hartford Region, School Year 2022-23, https://portal.ct.gov/-/media/SDE/School-Choice/RSCO/RSCOFamilyGuide.pdf.

[90] *See* DOE Production at 102 ("SDE will communicate information about the CAP framework for 2020-21 at a later date based, in part, on implementation of various initiatives in the Stipulation.").

C. Department of Administrative Services (DAS)

We submitted a FOI request to DAS on April 26, seeking information about its procurement, use, and assessment of algorithms in the administration of the hiring process for state employees and contractors. We were interested in this algorithm in view of concerns about how the use of such algorithms may generate biases in hiring.[91] Although the FOIA statute requires agencies to respond to FOI requests within four business days, we heard nothing from DAS for several months.

We wrote to DAS on September 4 to inquire about the status of our request, and a DAS FOI officer responded that she would investigate. After hearing nothing more, we followed up on September 15 to request an update. We received an update on September 17 from a DAS FOI officer who suggested a meeting "to narrow the scope of this request and to discuss the limits of our production." The response also warned that some of the requested information might be withheld under the trade secret exemption. Then began a weeks-long process of postponement and non-communication. For over a month we attempted each week to set up a meeting, proposing several dates and times, only to receive a response after the proposed meeting dates had all passed.

Eventually we got on the phone. On a call on October 20, the FOI officer informed us that DAS plans to withhold "anything substantive related to the actual algorithm," invoking the trade secret exemption as well as exemption 6 covering "[t]est questions, scoring keys, and other examination data used to administer a[n] . . . examination for employment," C.G.S.A. § 1-210(b)(6). DAS indicated it would produce the contract for the algorithm and the request for proposal that led to the contract after they have been reviewed to redact any proprietary information. As of the date of this report, two months later, we have yet to receive any documents.

Further, DAS's invocation of exemption 6 seems spurious given the nature of our requests. The purpose of exemption 6 is to prevent test takers from having access to test questions in advance, to prevent them from gaming the system. It is unclear why DAS concludes that disclosing information about the employment qualifications used by its algorithm to sort job applications could enable anyone to game the system. And if DAS really is relying on hiring criteria that it refuses to disclose, this raises obvious concerns about disparate impact.

---

[91] *See* Manish Raghavan & Solon Barocas, *Challenges for Mitigating Bias in Algorithmic Hiring*, BROOKINGS (Dec. 6, 2019), https://www.brookings.edu/research/challenges-for-mitigating-bias-in-algorithmic-hiring/.

Conclusions

Our requests to DCF, DOE, and DAS illustrate the many shortcomings with algorithmic accountability under current FOI laws, and suggest that new disclosure paradigms are needed for meaningful algorithmic accountability. The use of algorithms by the three agencies we tested remains opaque in several ways: the lack of information in the possession of the agency concerning the operation of algorithms used, any assessments of effectiveness or bias, and the manner of its procurement; FOI exemptions that preclude disclosure of information needed to understand and assess algorithms used to perform governmental functions; and, disclosure timelines that render after-the-fact FOI inquiries suboptimal as a means of ensuring algorithmic accountability. All three agencies indicated that their respective algorithms were proprietary and not subject to disclosure, depriving the public of the most crucial information for understanding the algorithm's logic. Moreover, the information we were able to obtain through the FOI requests suggests that the agencies themselves did not assess the algorithms for effectiveness or for bias, either before acquiring them or during their use.

PART 3: POSSIBLE LEGISLATIVE RESPONSES

Given the potential for government algorithms to inflict great harm, new methods for ensuring transparency and accountability are urgently needed. Further study is required to understand the full scope of the problem and to develop alternative ways for obtaining and making public essential information about the operation, impact, and effectiveness of government's algorithms.

In this final section, we canvas proposals for increasing algorithmic accountability that have been advanced by legislators and/or academics. Current proposals fall generally into three main buckets: (1) mandatory internal or external assessments of the algorithm's efficacy and potential for bias by qualified experts; (2) increased transparency, including through publication of any assessments of algorithms and/or trade secret reform to allow greater access to the workings of an algorithm; and (3) mandatory disclosures specifically to individuals affected by an algorithm. All are compatible with one another and worth considering as part of an overall legislative study of algorithmic accountability.

A. Mandatory Algorithmic Assessments

As shown above, there is reason to believe that governmental agencies are generally failing to assess whether the algorithms they use are ineffective or biased. This runs the risk—which has proven real too often—that citizens are subjected to error-prone decision-making systems or to unintended impacts across protected

characteristics like race or socioeconomic status. One obvious solution to this problem is to require assessments of an algorithm's efficacy and bias.

There are three dimensions to algorithmic assessments: (1) when the assessment is performed, (2) who performs it, and (3) whether the assessment has any follow-on effects. First, to be most effective, assessments would occur both *before* the agency acquires or use an algorithm (so as to identify any bugs or predictable adverse effects),[92] and at *regular intervals* while the algorithm is being deployed (so as to ensure that the algorithm is working as intended having no effects that were not predicted at the start, and properly accounting for new data and changed conditions). Second, maximizing the objectivity and quality of the assessments may require them to be performed by both the *agency* and *outside experts*. Third, the results of the assessment could affect who must approve the use of an algorithm or how it may be deployed.

One promising model for the first dimension, the "when," is found in a bill proposed in 2020 by U.S. Senator Sherrod Brown of Ohio called the Data Accountability and Transparency Act (the "DATA Act").[93] The DATA Act would have required public agencies to perform an "automated decision system risk assessment" before deploying an algorithm.[94] Requiring such an assessment would force agencies to evaluate the algorithm's development process, its design, and its training data "for potential risks to accuracy, bias, [and] discrimination."[95] The DATA Act also would have required agencies to perform an "automated decision system impact evaluation" on an annual basis.[96] These evaluations would entail an assessment of the algorithm's "accuracy, [and] bias on the basis of protected class" and the

---

[92] Pre-use impact assessments are sometimes likened to environmental impact assessments required by the National Environmental Protection Act, 42 U.S.C. § 4321, et seq. *See, e.g.*, Dillon Reisman et al., *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*, AI NOW INSTITUTE 7 (2018), https://ainowinstitute.org/aiareport2018.pdf.

[93] The text of the discussion draft of the bill is available at https://www.banking.senate.gov/imo/media/doc/Brown%20-%20DATA%202020%20Discussion%20Draft.pdf. Senator Brown's bill proposed comprehensive amendments to data privacy across the country, many of which are not directly relevant to algorithmic accountability for public agencies. We focus on the aspects of his proposal that bear directly on that issue.

[94] DATA Act § 105(c)(1). The DATA Act would have required this of each "data aggregator," which includes both agencies and private entities. *See* DATA Act § 3(8) (defining "data aggregator" to include "any person"); *id.* § 3(17) (defining "person" to include state agencies). A proposal for algorithmic accountability for public agencies could limit this requirement to public agencies.

[95] DATA Act § 3(6).

[96] DATA Act § 105(c)(2).

effectiveness of "measures taken to minimize risks as outlined in any prior automated decision system risk assessment."[97] They also would have required the agency to recommend measures to minimize risks to accuracy, or bias.[98]

A model for the second dimension, the "who," is Canada's Directive on Automated Decision-making (the "Directive"). The Directive requires public agencies to first perform an impact assessment that is standardized across all agencies. The assessment determines the algorithm's risk score. The lowest-risk algorithms do not require external evaluations, while higher-risk algorithms require different amounts of external peer review by defined individuals (for example, university faculty members, third-party vendors, or experts from non-governmental organizations) before the algorithm's deployment.[99] Another complementary option is to require or encourage prospective vendors to include an algorithmic assessment in their bid proposals, which the agency can then use to inform its own assessment.[100]

The Directive does not require ongoing peer review after the algorithm's implementation, but that need not be so. Others have recommended that algorithms regularly undergo external audits by outside experts to identify and recommend fixes for any kinks in their implementation. For instance, the Ada Lovelace Institute, a British research institute focusing on the just and equitable use of artificial intelligence, recommends "bias audits" conducted by external experts. In principle bias audits can be conducted with or without access to the code of the system. For example, one could audit hiring algorithms by "participating in them," such as by submitting identical job applications but varying the applicants' race.[101] Effectively auditing machine learning algorithms that update over time, such as the Mindshare

---

[97] DATA Act § 3(5).

[98] DATA Act § 3(5).

[99] *See* Directive on Automated Decision-making, § 6.1 & App. C (2019), https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592.

[100] This is roughly the approach recommended by pending proposed legislation in California called the Automated Decision Systems Accountability Act ("ADSAA"), whose current text is available at https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202120220AB13.

[101] Ada Lovelace Institute, *Examining the Black Box: Tools for Assessing Algorithmic Systems*, 9 (2020). This report can be downloaded from the following link: https://www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems/.

algorithm discussed in Parts 1 and 2, may require legislation mandating that agencies save the code of an algorithm at regular intervals.[102]

Canada's Directive also provides a model for the third dimension of algorithmic assessments, namely what the agency must do in response to them. This, too, depends on the level of risk. For instance, the Directive requires humans to be involved in decisions made by high-risk algorithms, requires more and better-documented training for agency staff using high-risk algorithms, and requires more senior approval the higher the algorithm's risk score.[103]

B. Increased Transparency

Transparency is another dimension of algorithmic accountability and, as discussed in Part 2, was sorely lacking in our limited Connecticut experiment. We discuss two broad ways to increase transparency: affirmative disclosure and amendments to trade secret laws.

First, a legislature could require agencies to affirmatively disclose (including on their websites) certain information about their use of algorithms. This information should include, at a minimum, a description of each automated decision system used by the agency and all risk assessments and impact evaluations performed by the agency, as many legislative proposals would require.[104] It should also include the agency's plans to acquire an algorithm with sufficient time for public comment, which, as shown in Parts 1 and 2, is often opaque when algorithms are acquired outside the ordinary procurement process. It may also require disclosure of any source code developed by the agency and not otherwise subject to trade secret protection—the Directive, for example, generally requires agencies to disclose the source code of algorithms that they own.[105]

While there are always tradeoffs between affirmative disclosure and request-based systems,[106] affirmative disclosure may be especially appropriate here.

---

[102] For example, one scholar recommends that agencies record "a comprehensive history of decisions made in a case," including "the actual rules applied in every mini-decision that the system makes." Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1305 (2008), https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=1166&context=law_lawreview.

[103] *See* Directive, App. C.

[104] *See, e.g.*, Directive, § 6.1.4; DATA Act, § 105(d); ADSAA, § 12116.5(a). Of course, the agencies could redact information that is exempt under the Freedom of Information Act.

[105] *See* Directive, § 6.2.6.

[106] *See, e.g.*, David E. Pozen, *Freedom of Information Beyond the Freedom of Information Act*, 165 U. PENN. L. REV. 1097, 1148-55 (2017),

Algorithms are generally less conspicuous than most government programs. This means potential FOIA requesters may often have no idea they exist, and hence have no reason to request information about them. Moreover, algorithms will often affect large and diverse groups of stakeholders who may not share the results of their FOIA requests. Plus, the length of time agencies often take to respond to FOIA requests—as evidenced by our experience, discussed in Part 2—may well obviate the relevance of any information the requester ultimately obtains. If, for instance, agencies assess their algorithms on an annual basis, the assessment the requester obtains might no longer be relevant. And requiring agencies to affirmatively disclose basic information about the algorithm and its assessment would not be resource intensive.

Second, another way to increase algorithmic transparency is to remove or limit the trade secret protection available for algorithms, at least in certain cases. As shown above, trade secret laws impede the public from assessing the effectiveness and potential bias of algorithms used by public agencies but bought from private parties. While there are countervailing factors favoring trade secrecy, there is no clear reason why those factors should always triumph over the public's need to ensure its agencies' systems are working properly and equitably.[107] The legislature could balance these factors by requiring private parties to disclose the source code and (suitably anonymized) training data only for algorithms that exceed a certain risk threshold. Risk could be determined either by the agencies, through a Directive-style impact assessment, or by the legislature, through defining certain classes of algorithms it expects in advance to be high-risk. Idaho has recently taken the latter approach for algorithms used in pretrial risk assessments, requiring owners of those algorithms to waive trade secret protection.[108]

As an alternative, the law could require the algorithm's owner to grant the agency (and perhaps also experts retained by the agency) full access to the algorithm as a condition of acquisition—an alternative whose efficacy could be bolstered by requiring impact assessments. The Directive takes this approach.[109] It is clear from our

---

https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=9579&context=penn_law_review (discussing tradeoffs and advantages of affirmative disclosure).

[107] *See generally* David S. Levine, *Secrecy and Unaccountability: Trade Secrets in Our Public Infrastructure*, 59 FLA. L. REV. 135 (2007), http://www.floridalawreview.com/2010/david-s-levine-secrecy-and-unaccountability-trade-secrets-in-our-public-infrastructure/ (arguing that the application of trade secrets law to public infrastructure is inconsistent with democratic accountability).

[108] *See* Idaho Code § 19-1910.

[109] *See* Directive, § 6.2.5.

request to DCF that Connecticut agencies at least sometimes license algorithms without insisting that they have access to them, which prevents them from independently assessing the algorithm for efficacy or bias. This degree of transparency would cause no undue harm to the algorithm owner's commercial interests but would have a significant positive impact on democratic accountability.

C. Disclosure to Individuals

The accountability measures discussed above will improve the public's ability to understand whether algorithms in general are biased or ineffective. But these holistic assessments will not disclose specific instances in which an algorithm might have worked improperly—for instance, they might not reveal if a particular person was erroneously denied a benefit to which he or she was entitled. Mitigating that problem requires empowering individuals to ascertain whether an algorithm has mistreated them and take corrective action if necessary. The European General Data Protection Regulation (GDPR) provides one such model.[110]

For instance, the GDPR requires data controllers—including public agencies—to inform data subjects, upon the collection of their data, whether the data will be subjected to processing by an algorithm and, if so, approximately how the algorithm works and the significance and consequences of the data processing. *See* GDPR Art. 13(2)(f). This is a form of affirmative disclosure regime on an individual level. For example, rather than (or in addition to) disclosing its use of a child welfare algorithm on its website, DCF could inform a family when a case is opened that it may be subject to the Mindshare algorithm.

Under the GDPR, armed with the knowledge that his or her data will be processed by an algorithm, the data subject may then access the data to ensure its accuracy, *see* GDPR Art. 15(1), and correct any inaccuracies, *see* GDPR Art. 16. These measures allow individuals to monitor how agencies' algorithms use their data and ensure they do so correctly.[111]

CONCLUSIONS

Government transparency and accountability laws must reflect what government is doing. More and more, governments are making decisions by

[110] The GDPR makes exceptions for law enforcement and other activities that require data processing to be kept secret from the data subject. *See* GDPR Art. 2(2).

[111] The GDPR addresses data protection more generally, and thus its rights apply beyond algorithmic decision-making. The Legislature could limit the scope of those rights to address algorithm-specific issues.

algorithm. As this study has shown, decisions by algorithm raise issues of potential ineffectiveness and bias for which public oversight is crucial. Yet the law has not kept pace with those issues, creating dangerous gaps for public participation in government. This paper has outlined some possible measures to fill those gaps. These measures, and the underlying problem, require serious legislative consideration and study.