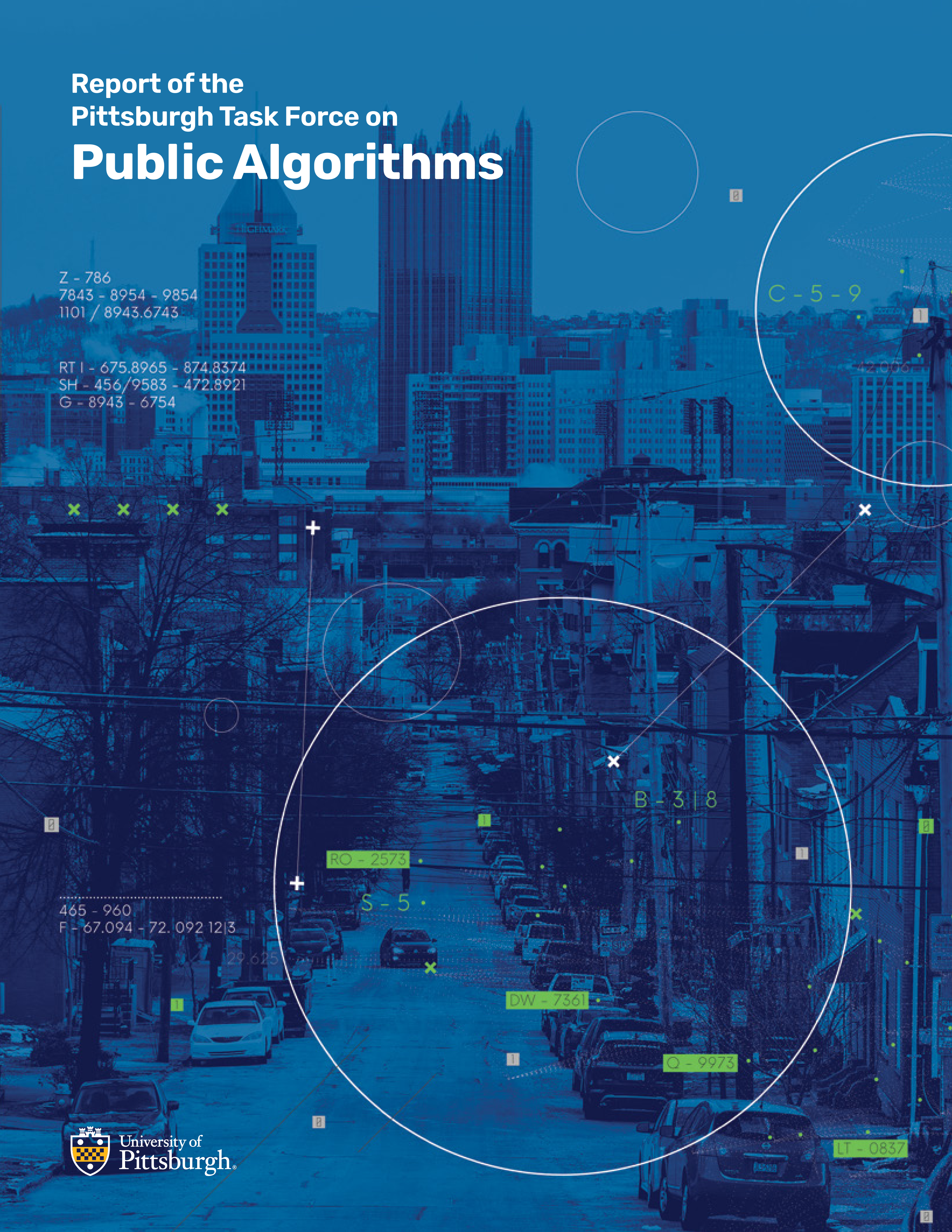
Report of the
Pittsburgh Task Force on
Public Algorithms



University of Pittsburgh

Contents

Task Force Members 2

Acknowledgments 4

Introduction 5

Executive Summary 6

I. What is a Public Algorithmic System? 8

II. Public Algorithmic Systems in the Pittsburgh Region 9

III. Potential Problems with Public Algorithmic Systems 14

IV. Task Force Recommendations 21

V. Additional Best Practices for Municipal Governments 28

VI. Conclusion 31

Appendix: Community Feedback Summary 32

Task Force Members

MEMBERS OF THE PITTSBURGH TASK FORCE ON PUBLIC ALGORITHMS

Rachel Berger, MD, MPH – chief, Child Advocacy Center, UPMC Children's Hospital of Pittsburgh; director, child abuse research, Safar Center for Resuscitation Research, University of Pittsburgh; professor of pediatrics, University of Pittsburgh School of Medicine

Esther Bush – former president, Urban League of Greater Pittsburgh

Alexandra Chouldechova – Estella Loomis McCandless Associate Professor of Statistics and Public Policy, Carnegie Mellon University

David Danks – professor of data science and philosophy, University of California, San Diego

Christopher Deluzio – policy director, Pitt Cyber, University of Pittsburgh

Ángel Díaz – lecturer in law, UCLA School of Law

Courtney Ehrlichman – chief executive officer, the Ehrlichman Group

Rosta Farzan – associate dean for diversity, equity and inclusion, associate professor, School of Computing and Information, University of Pittsburgh

Sara Goodkind – associate professor, School of Social Work, Department of Sociology and Gender, Sexuality, and Women's Studies Program, University of Pittsburgh

Ellen Goodman – professor of law, Rutgers University School of Law

David Harris – Sally Ann Semenko Endowed Chair, professor of law, University of Pittsburgh School of Law

David Hickton – founding director, Pitt Cyber, University of Pittsburgh

Yu-Ru Lin – associate professor, School of Computing and Information, University of Pittsburgh

Sharon McDaniel – founder, president, and chief executive officer, A Second Chance, Inc.

Michelle McMurray – vice president of program and community engagement, the Pittsburgh Foundation

Mark A. Nordenberg – chair, Institute of Politics, University of Pittsburgh; chancellor emeritus of the University; distinguished service professor of law

Beth Schwanke – executive director, Pitt Cyber, University of Pittsburgh

LaTrenda Sherrill – lead consultant, Common Cause Consultants; former deputy chief (education), Office of the Mayor of the City of Pittsburgh

Michael Skirpan – executive director, Community Forge; special faculty, Department of Philosophy, Dietrich College of Humanities and Social Sciences, Carnegie Mellon University

Nyssa Taylor – criminal justice policy counsel, ACLU of Pennsylvania

Frederick Thieman – Henry Buhl Jr. Chair for Civic Leadership, Buhl Foundation; chair, Community College of Allegheny County Board of Trustees; former U.S. attorney for the Western District of Pennsylvania

Task force members are serving in their individual capacity.

GOVERNMENT ADVISORY PANEL

Chris Belasco – manager, data services, Department of Innovation and Performance, City of Pittsburgh

Erin Dalton – director, Allegheny County Department of Human Services

T. Matthew Dugan – chief public defender, Allegheny County Office of the Public Defender

Joanne Foerster – director of performance and analytics, Allegheny County

Heath Johnson – manager, crime analysis, Pittsburgh Bureau of Police

Laura Meixell – enterprise data architect, Allegheny County Department of Human Services

Shatara Murphy – assistant director, community affairs, Department of Public Safety, City of Pittsburgh

The task force was served by an advisory panel composed of designees from Allegheny County and the City of Pittsburgh. These members and their respective offices are not signatories to the task force's final report and served in an advisory role only.

Acknowledgments

The University of Pittsburgh Institute for Cyber Law, Policy, and Security (Pitt Cyber) and the Pittsburgh Task Force on Public Algorithms gratefully acknowledge the generous support of The Heinz Endowments and the Hillman Foundation.

The task force thanks the entire staff of Pitt Cyber, including Sarah Barca and Kate Ulreich, for their many contributions to this work.

Many thanks to those who provided thoughtful feedback to the task force as it drafted this report: Eric Bloedorn, chief scientist for artificial intelligence (AI) adoption, the MITRE Corporation; Gretchen Greene, fellow, Harvard Kennedy School, senior advisor, the Hastings Center; Jonathan Rotner, human-centered technologist, the MITRE Corporation; and Vincent Southerland, assistant professor of clinical law, co-faculty director, Center on Race, Inequality, and the Law, New York University School of Law. Thanks to Megan Turnbull and Weiss Burkardt Kramer, LLC, for legal support and counsel.

We are grateful for the advice and guidance from local government personnel who served on the government advisory panel (members listed on page 3), as well as Nick Hall and James O'Toole for their prior service on the government advisory panel.

We acknowledge the thoughtful contributions we received in public comments and through our community engagement, which task force member LaTrenda Sherrill expertly led, and the many organizations and people throughout Pennsylvania and the country who provided us with their expertise and insights. In particular, thanks to those who presented at the task force's meetings: Janice Dean, director of Allegheny County Pretrial Services; Sylvia Harris, assistant director of operations, Department of Innovation and Performance, City of Pittsburgh; Dr. Heath Johnson, crime analysis coordinator, Pittsburgh Bureau of Police; Ellen Goodman, professor, Rutgers Law School; Cathy O'Neil, Tom Adams, and Jake Appel, ORCAA; Rashida Richardson, assistant professor of law and political science, Northeastern University School of Law; LaTrenda Sherrill, Common Cause Consultants; Rhema Vaithianathan, professor of economics and codirector of the Centre for Social Data Analytics, Auckland University of Technology; and Meredith Whitaker, co-founder, AI Now.

We also express gratitude to the following people, who provided thoughtful advice, support, and assistance in many different ways that ultimately shaped the task force's work: Kenny Chen, Mike Colaresi, Kara Dally, Carl DiSalvo, Daren Ellerbee and the University of Pittsburgh Community Engagement Center in Homewood, David Frisch, Cherylie Fuller, Bob Gradeck and the Western Pennsylvania Regional Data Center, Sera Linardi, Lance Lindauer, Brace Lowe, Michael Madison, Breanna McVay, Veronica Oviedo, Alka Patel, Hannah Sassaman, John Wallace, Theodore Wilhite, Aimee Wilkins, and many others.

Most importantly, we thank the many community members who generously shared their time, experiences, and expertise with the task force.

Introduction

The Pittsburgh Task Force on Public Algorithms came together in 2020 to study the use of algorithms in local government. The task force is an independent body, hosted by the University of Pittsburgh's Institute for Cyber Law, Policy, and Security (Pitt Cyber). We were assisted by a government advisory panel, whose members served only in an advisory role and thus are not signatories to this report.

We approached this work with three core understandings. First, government use of algorithms can bring many concrete benefits. Second, those same algorithms can carry risks that can and should be guarded against. And, finally, government acceptance and use of algorithms will likely increase in the coming years.

With these understandings in mind, we offer recommendations and best practices for the Pittsburgh region's governments to consider. We believe that this work offers approaches that facilitate the transparent use of algorithms by our local governments while still providing needed external guardrails to ensure that our algorithmic systems do not lock in and exacerbate bias or other harms. We aimed to facilitate a balance of mitigating potential for harm, creating an environment for more equitable use of algorithms, and ensuring valuable innovation for the good of the region's residents.

Across the country, government agencies have a mixed track record of employing algorithms. Some have used algorithms to improve on prior human-based processes and outcomes and have done so with transparency, humility, and community participation. Some have not. However, we have not provided an evaluation of individual agency use of algorithms. Instead, this report looks to deliver concrete and achievable recommendations that, if adopted, could serve to guide the appropriate development and use of algorithms by governments going forward. Where local government actors have embraced transparency and equity in the development of algorithms, the task force looked to build upon that foundation in developing our recommendations.

There is a growing realization that algorithms can—and in some cases do—have extraordinary power to shape our lives, in both the private and public sectors. We believe that this is a critical moment for our region's governments to act. The recommendations in this report provide a meaningful path to better serve our residents and to lead as a model for municipalities across the country in responsible use of government algorithms.

Executive Summary

Algorithms are already here in the Pittsburgh region's governments. The systems are poised to continue their growth, offering the promise of efficiency and improved decision-making.

In the Pittsburgh region and beyond, algorithms are increasingly driving and shaping governmental decisions with tremendous impact on our lives. Child welfare investigations, bail determinations, and a variety of other public functions more and more depend on human-trained machines' ability to guide outcomes. Section II of this report provides an overview of two illustrative algorithmic systems in our region.

And yet, despite their increasing prevalence, the public often knows little about government algorithmic systems: their goals, how they work, who designed them, and more. Some agencies have endeavored to bring in the public, whereas others have not.

Government agencies should want public participation and understanding of these growing and powerful tools that harness extraordinary amounts of data. At present, little requires our regional governmental agencies to share information about algorithmic systems or to submit those systems to outside and public scrutiny.[1] Moreover, there is evidence that some algorithmic systems can lock in and exacerbate bias and harms (especially along racial and gender lines), leading to more inequity and injustice.[2]

Algorithmic systems can, however, offer significant benefits to the public: more efficient processing of data, fewer errors in decision-making relative to humans or perhaps even less biased decision-making, and the ability to consider vast troves of factors and data. There are government functions where algorithmic systems can be an improvement over human effort and/or can effectively supplement human effort to bring better results to our region.

Against this complex backdrop, the task force endeavored to learn from our local governments' experiences with these systems. We observed a range of approaches with profoundly different commitments to being transparent, engaging the public, and obtaining outside reviews of systems.

We also sought to listen to residents, especially those in communities most likely to be affected by governmental algorithms. Although the COVID-19 pandemic complicated that effort, some themes emerged: a sense of being kept in the dark, frustration that government is harnessing data to target enforcement instead of deliver resources, and a desire for more democratic deliberation and transparency regarding these systems. But there was also an appreciation of the promise of algorithms.

Within this context, the Pittsburgh Task Force on Public Algorithms offers concrete recommendations to manage the accountability of public algorithmic systems in our region. We do so with humility, but we hope that these recommendations, if implemented, will offer transparency into government algorithmic systems, facilitate public participation in the development of such systems, empower outside scrutiny of agency systems, and create an environment where appropriate systems can responsibly flourish.

There is no direct model for this work, although there are increasing experiments in algorithmic accountability efforts at all levels of government across the globe. Where we could, the task force looked to examples on which we could rely in crafting our recommendations, with particular attention to local government actors (such as Allegheny County's Department of Human Services) that have a track record of developing algorithmic tools in public view.

Section IV of the report explains the task force's recommendations in detail. In sum, our top-line recommendations for our region's governments are:

- Encourage meaningful public participation, commensurate with the risk level of the potential system.
- Involve the public in algorithmic system development plans, from the earliest stages through any later substantive changes to the system.
- Utilize third-party reviews when the system might be higher risk.
- Procurement and contracting (including data-sharing agreements) processes could include review to assess whether any planned procurement might include an algorithmic system.
- Publish information about algorithmic systems on a public registry website.
- Avoid facial-recognition and related systems.
- Evaluate effectiveness of recommendations.

There is a tremendous opportunity for the region to fashion a framework for managing and harnessing public algorithmic systems. Such actions will require investments, ones that the task force believes are worthy of support. If we are successful, municipalities across the country could look to this framework as a model, positioning Pittsburgh as a leader and helping to achieve greater algorithmic justice and accountability balanced with responsible growth.

I. What is a Public Algorithmic System?

Throughout this report, the task force principally uses the term *public algorithmic system*, defined in full as follows:

> The term **"public algorithmic system"** means any system, software, or process that uses computation, including those derived from machine learning or other data-processing or artificial intelligence (AI) techniques, to aid or replace government decisions, judgments, and/or policy implementations that impact opportunities, access, liberties, rights, and/or safety.[3]

Terms like "algorithms," "artificial intelligence," "automated decision system," "predictive risk model," and "machine learning" tend to proliferate in any discussion of governments using big data in decision-making. Many of these terms have substantial overlap and, in common usage, are used imprecisely. They are not perfectly interchangeable, of course, and nuances in meaning can have important distinctions in understanding the tool in question.[4] For the purposes of the task force's recommendations, however, these nuances matter less than an understanding that governments increasingly use these tools to make or help make decisions.

The task force's definition, which is heavily informed by Professor Rashida Richardson's work[5], is broader than the usual definition of "algorithms," which are "generally regarded as the mathematical logic behind any type of system that performs tasks or makes decisions"[6] or "specific sequences of steps used to accomplish some task, especially those embedded in a computer."[7] In other words, "algorithms" are typically understood as problem-solving functions rather than the complete system that might encompass an algorithmic function. For our purposes, however, this narrow understanding of algorithms would be too limited a view. Indeed, as one study has observed, "[T]oday's automated decisions are not defined by algorithms alone. Rather, they emerge from *automated systems* that mix human judgment, conventional software, and statistical models, all designed to serve human goals and purposes."[8]

The potentially broad nature of this definition could lead to capturing a wider set of systems than intended.[9] However, as described more fully in the forthcoming recommendations of this report, the task force believes that the relevant risk of any given system should determine the level of concern and corresponding scrutiny for that system. In this framework, the risk-based approach that we have taken should mitigate overbreadth concerns.

Critically, the task force and its work are focused on *public* algorithmic systems, meaning those systems that governments use (including when a government actor relies on or partners with a private algorithmic system). This excludes non-governmental systems, such as those that a prospective private employer, a social media company, or a private health insurer might use. Moreover, the task force's work is focused on algorithms in government decision-making, which, irrespective of whether it might be shaped by an algorithm, can carry risks, especially when lacking public input and involvement. Many of the task force's suggestions about bolstering public participation in algorithmic system development could similarly foster trust in government decision-making, more broadly, though this is beyond the scope of our work.

II. Public Algorithmic Systems in the Pittsburgh Region

Our region has been at the forefront of applying algorithmic systems to governmental functions.

The following sections highlight two key systems: (1) the Allegheny County Department of Human Services (DHS) child welfare tool, the Allegheny Family Screening Tool (AFST), in light of its prominence; the county's transparency around its development, deployment, and use; and the county's efforts to secure external evaluations and public feedback, and (2) the City of Pittsburgh's suspended predictive policing system, which, like in most government agencies across the country, lacked such process measures. In developing the recommendations in this report, the task force paid particular attention to the DHS experience with AFST and efforts to incorporate transparency and accountability into the system's development. The task force was impressed with DHS's goal of a more open and consultative process, something that has not been the norm for most government agencies across the country.

The two systems discussed below—and the other systems identified—are not an exhaustive list of systems in use in our region; rather, they are illustrative of significant ways that our region's governments are turning to algorithmic systems to aid in governmental decision-making and the processes around their adoption and use. The task force *did not* endeavor to audit or scrutinize the workings of any specific system; instead, we focused on the processes surrounding system development and implementation to inform our recommendations.

The task force anticipates that the growth of these algorithmic tools is poised to continue in the Pittsburgh region and beyond, bolstered by the availability of large troves of data, the power and promise of algorithmic systems, and the recent experience of local governmental officials in developing and deploying such systems.

ALLEGHENY FAMILY SCREENING TOOL (AFST)

The AFST is an algorithmic system[10] that helps child-welfare personnel assess whether or not to "screen in" a referral for an investigation when they receive an allegation of certain types of child maltreatment (e.g., neglect). Allegheny County DHS developed the system, in use since August 2016, with a team from Auckland University of Technology.

The process before adoption of the AFST required staff at a child-abuse call center receiving referrals—that is, allegations of maltreatment—"to manually access a myriad of data and information to help decide whether or not to investigate the allegation ('screen in' and investigate or 'screen out' and offer relevant community resources)."[11] In essence, staff had to comb through available data and then make judgment calls about whether an investigation was warranted.

Much of the motivation behind developing the system was to better inform such decisions made at the critical time when allegations are received, when there is often limited information received from callers on which to base decisions, and to "reduce variability in staff decision-making."[12] According to DHS, before the AFST, "[A]n analysis found that 27% of highest risk cases were being screened out and 48% of the lowest risk cases were being screened in."[13] As with nearly all child-welfare departments, Child Protective Services does not have the resources to investigate all referrals.

The AFST therefore looks to help determine which cases are most critical to investigate by augmenting the manual decision-making process. The AFST generates an algorithmically indicated risk score, which is then reviewed by a screener, who decides whether or not to screen in a referral for an investigation. The AFST makes an adverse outcome prediction—that is, whether a child will experience a placement within two years of the screening.[14] Placement refers to whether a child will be removed from the current home situation by DHS and placed in a different environment. According to DHS's AFST "frequently asked questions" document, "Only the call screener and call screening supervisor have access to the AFST score," and the investigations staff and the courts do not have access to the AFST score.[15] In other words, no individual who participates in decisions about whether children should be placed in different environments currently has access to scores.

Underlying the AFST is the DHS "Data Warehouse." The county describes the Data Warehouse as "bring[ing] together and integrat[ing] client and service data from a wide variety of sources both internal and external to the County. It was created by consolidating publicly-funded human services data (e.g., behavioral health, child welfare, intellectual disability, homelessness and aging) and, over time, expanded to include data from other sources."[16] The AFST relies on a subset of data sources from the Data Warehouse: child-welfare records, jail records, juvenile-probation records, behavioral-health records, and birth records.[17] A DHS report on the AFST's methodology includes in an appendix the "full list of predictor fields used in" the latest version of the AFST.[17] DHS intends to incorporate additional data into AFST in the future—including data from private insurers—with an eye toward making the data more representative.[18]

///////////////

"Since August 2016, the Allegheny County Department of Human Services (DHS) has used the Allegheny Family Screening Tool (AFST) to enhance our child welfare call screening decision making process with the singular goal of improving child safety.

The AFST is a predictive risk modeling tool that rapidly integrates and analyzes hundreds of data elements for each person involved in an allegation of child maltreatment. The tool can rapidly integrate and analyze these data, housed in the DHS Data Warehouse, and create a synthesized visualization of the information. The result is a 'Family Screening Score' that predicts the long-term likelihood of future involvement in child welfare. By combining the insight gained through the score with other traditionally gathered information, a better prediction can be made of the long-term likelihood that the child will need to be removed from the home in the future.

According to the algorithm, the higher the score, the greater the chance of future out-of-home placement. When the score is at the highest levels, meeting the threshold for 'mandatory screen in,' the allegations in a call must be investigated. In all other circumstances, the information summarized by the score does not replace clinical judgment but rather provides additional information to assist in the call screening decision making process. The Family Screening Score is not used to make investigative or other child welfare decisions and is not shared beyond call screening."

Source: Allegheny County Family Screening Tool.

EVALUATIONS AND TRANSPARENCY

DHS solicited proposals "to design and implement Decision Support Tools and Predictive Analytics in Human Services" in February 2014.[19] Supported by both private philanthropic and government funds, DHS ultimately awarded a contract to a team from Auckland University of Technology, led by the codirector of the Centre for Social Data Analytics, Rhema Vaithianathan.[20] An external member of the vendor-selection committee, who is also a member of the task force, told us that she looked to "center the selection based on race, bias, and DHS's history. This early acknowledgement of bias and the attempts to mitigate the same in the early design was critically important." DHS also subjected the AFST to an external ethical analysis,[21] to which DHS provided a fulsome response.[22]

DHS sought to involve members of the region in the decision to use the AFST. As part of the development, community and stakeholder meetings were held. Calling such efforts "a priority for the County throughout the project," DHS held a variety of meetings and undertook efforts to bring in the public:

> "The County sought input from the community through various meetings, including six project-specific meetings. Three were held at early stages of the project to collect feedback from key external stakeholders and funders. DHS then held three open community meetings where over 30 stakeholder groups (including the Courts and the ACLU) were invited to discuss the work to date, implementation timeline and results. Additionally, DHS shared project updates with existing community networks including the Children's Cabinet and the Children, Youth and Families Advisory Board, and through the DHS Speaker Series."[23]

Separate from, and concurrent with, the development of the AFST itself, DHS also sought two related evaluations of the AFST (in addition to the ethical analysis noted above): a process evaluation and an impact evaluation. Hornby Zeller Associates conducted the process evaluation[24], and Stanford University performed the impact evaluation.[25] Critically, the impact evaluation determined that, among other things, the "AFST led to reductions in disparities of case opening rates between black and white children," an important improvement on the higher case-opening rates for Black children pre-AFST.[26] The evaluation also found increased "identification of children determined to be in need of further child welfare intervention" and that the AFST "did not lead to decreases in re-referral rates for children screened-out without investigation."[27] DHS also conducted external validation of the AFST, using data from UPMC Children's Hospital of Pittsburgh, demonstrating that the children whom the AFST identified as at highest risk were also those more likely to have hospital visits for treatment suggestive of maltreatment.[28]

PREDICTIVE POLICING SYSTEM[29]

Beginning in 2017, the Pittsburgh Bureau of Police and the Department of Innovation and Performance experimented with a predictive policing algorithmic system in partnership with Carnegie Mellon University's Metro21 (Smart Cities Institute). That system is currently suspended—something that the task force urged the city to do.[30] In responding to the task force's request that the city keep the program suspended, then-Mayor Bill Peduto wrote that he and other public safety leaders "share your concerns about the potential for predictive policing programs to exacerbate implicit bias and racial inequities."[31]

Despite that suspension, the task force included discussion of the system to highlight how opaque its development and operation were, reinforcing the need for meaningful public involvement in algorithmic system development in our region.

Of course, the public might weigh the expected benefits and risks of such a system and ultimately support it—but that is impossible without insights into the system, its workings, its data, and avenues for participation. Several members of the task force found the initial reported results of the system compelling (and a likely improvement over the legacy, human-based decision-making that the system was complementing), whereas others had significant concerns about the data used or related issues. All, however, agreed that the lack of transparency and public accountability justified the program's suspension. Such opacity seems to be far too common across the country, especially in policing applications of algorithmic systems.

Before the program's suspension, the system was implemented city-wide (in both residential areas and business districts). The funding for the partnership with Metro21 ended at the end of 2019. According to a since-removed posting on the Metro21 website:

> ///////////////
>
> "The team developed a predictive analytics program and policing strategy that is being used by the Pittsburgh Bureau of Police to predict crime hot spots and prevent serious violent crimes. A neural network model predicts locations that will likely have crime flare-ups in the following week. Police then use community policing approaches to patrol and deter crime in the hot spots with the aid of real-time crime mapping from their patrol car laptop computers."
>
> Source: Wil Gorr and Daniel O'Neill, Pittsburgh Crime Hot Spot Program: Preventing Crime with Predictive Policing, Carnegie Mellon University Metro21: Smart Cities Institute, https://www.cmu.edu/metro21/projects/reducing-crime.html (accessed November 2019).

Researchers behind the project, which began in 2017, made available publicly a paper that provides insights into how the project was developed and had been working in practice. The initial research project intended to assess the impact of predictive policing strategies on crime.

The subsequently developed system relied on historical reports of crimes and 911 calls to make predictions about where crime might occur in the future. Based on the system's predictions, Pittsburgh Police then pursued a "'hot-spot-based' approach": "Pittsburgh's hot spots are 500 sq. ft. areas where crime is likely to occur, and they can be either 'chronic,' meaning crime is likely to happen there at all times, or 'temporary,' meaning they have occasional spikes of crime activity."[32] The police targeted hot spots for "treatments"—that is, deployment of police based on the algorithmic system's predictions.[33] According to the Bureau of Police, before the program's suspension, each police zone had been receiving six hot spots for patrols (three chronic, three temporary). According to the researchers, they found a 25.3% reduction in serious violent crimes per hot spot (relative to the control hot spots).[34]

The project apparently relied on two data sets from a five-year period from 2011 to 2016: (1) "all 206,150 crime incidents reported by the [Pittsburgh Bureau of Police]," which were compiled from Pittsburgh's Automated Police Reporting System (APRS), and (2) "information on 911 calls for assistance to the Pittsburgh police, totaling approximately one million calls over the five-year period of analysis."[35] Of course, crimes that go unreported are not included in these data.

As *PostIndustrial* reported in June 2019, the system's data "include[d] three broad sources: past crimes of the same kind as the target crime being predicted, other kinds of past crimes which are found to be predictive of the target crime, and 911 calls about crimes."[36] The Bureau of Police advised the task force that personally identifiable information, demographic data, and officer-initiated actions (e.g., traffic stops or arrests) were not included in the system's data.

EVALUATIONS AND TRANSPARENCY

The task force is not aware of any external review of the program,[37] nor did the task force identify any effort to invite the public to participate in the design of the program, to solicit opinions about its development or implementation, or to otherwise educate the public about the system and its effects before the system went into effect. In March 2021, one of the program's researchers submitted a comment to the task force, sharing access to other documents about the system.[38]

As with any program relying on historical crime-reporting data—which can reflect long-standing racial and economic disparities in Pittsburgh—this predictive policing system involves sensitive issues of policing (including the critical issue of where to send police for potential enforcement) with implications for public safety and fairness.

OTHER SYSTEMS

There are other instances of algorithmic systems in use in our region. County DHS has developed a system designed to better match residents and homelessness resources in Allegheny County[39] and has introduced a program called Hello Baby with the goal of effectively and efficiently getting available resources to infants and families.[40] As with the AFST, county DHS has made several reports available, with others to come once completed: a methodology report,[41] two ethical reviews,[42] and the DHS response to those ethical reviews.[43]

City algorithmic systems have included those to manage traffic lights,[44] a system that detects and locates gunshots,[45] and a system that predicts risk scores for fires at commercial properties.[46] In addition, the Pittsburgh Water and Sewer Authority is endeavoring to use machine learning to predict the location of lead-contaminated service lines,[47] and Pittsburgh Public Schools have harnessed data to flag whether students might be at risk of falling behind academically and to evaluate teacher performance.

As with many jurisdictions across the country, the judiciary also uses algorithmic systems. Since 2007, Allegheny County's courts have used an internally developed pretrial risk-assessment tool to aid judicial decisions about whether to confine criminal defendants before trial. Allegheny County courts also have access to the Public Safety Assessment, developed by Arnold Ventures,[48] which the courts implemented in 2016 in parallel to the existing risk-assessment tool, and juvenile detention decisions are shaped by the Pennsylvania Detention Risk Assessment Instrument.[49]

Elsewhere in the region, researchers from the University of Pittsburgh have been exploring a system designed to predict opioid overdoses and "to identify individual risk factors,"[50] with likely application in local governments.

III. Potential Problems with Public Algorithmic Systems

The task force believes that government consideration of any algorithmic system should entail (1) a comparison of that system (and its expected benefits) to the existing human decision-making process it would supplant or supplement and (2) a weighing of those benefits against expected or potential problems (and mitigations) inherent in the system or its data. A weighing of these potential problems against expected benefits and improvements over existing human decision-making might lead policymakers to proceed with adopting algorithmic systems for some applications—but perhaps rejecting them in others. Such balancing of costs and benefits, however, must be done with a full appreciation of all perspectives and should include robust public involvement to improve outcomes and to engender trust.

Some algorithmic systems have the potential to perform complex tasks more efficiently and fairly than humans.[51] Proponents of these tools also point to the potential for reducing inaccuracies, biases, and other common shortcomings of human decision-making, and they even argue, "[W]hen algorithms are involved, proving discrimination will be easier—or at least it should be, and can be made to be."[52] As a leading management consultancy that often advises governmental clients puts it, these tools can "enable governments to perform more efficiently, both improving outcomes and keeping costs down."[53] Some argue that algorithmic systems have "enormous potential for good," and "could greatly improve quality of life and help individuals meet long term goals."[54]

It is no surprise then, that such public systems are growing in prevalence. Indeed, "[A]n increasing number of government agencies are considering or starting to use [artificial intelligence and machine learning] to improve decision making."[55] The task force is not aware of any comprehensive and complete accounting of the numbers of automated decision-making systems currently in use in municipalities but anticipates that the numbers and impact will become even more significant in the next decade, including in our region.

Yet governmental use of algorithmic systems—like any type of decision-making—can indeed suffer from inaccuracies, misuse, and bias, as well as a host of other problems. Some of these problems are unique to algorithmic systems and the data on which they rely. This section will survey some of the common problems implicated by government algorithmic systems.

ILLUSTRATIVE PROBLEMS

Algorithmic systems—like the human decision-making processes they might supplant or aid—can be problematic. Algorithmic systems can suffer from high error rates and can lead to inequitable and biased outcomes if the underlying data are similarly flawed. Moreover, issues with training, designing, and implementing algorithmic systems can lead to feedback loops—where outputs of algorithmic systems go on to affect the data inputs and thus subsequent outputs.[56]

Consider this cautionary tale from Michigan of a public algorithmic systems gone awry. The state's Unemployment Insurance Agency turned to outside vendors to automate processes regarding unemployment data and to ultimately adjudicate and impose penalties for benefits fraud.[57] The algorithmic systems replaced humans who had been involved in various aspects of scrutinizing unemployment claims and was programmed or trained in a way that made it quite likely to falsely flag instances of fraud. Official public details about the cause of these false positives are scant, but, according to a report from AI Now Institute, the system "falsely identified more than 40,000 Michigan residents of suspected fraud" from October 2013 to August 2015.[58] In essence, this Michigan algorithmic system, which was intended to identify unemployment fraud, incorrectly flagged tens of thousands of Michiganders for suspected fraud with significant consequences for those flagged, including civil penalties, wage garnishments, and tax-refund seizures.[59]

Not all algorithmic systems suffer from such obvious (or detectable) errors, and there are a variety of flaws that can imperil algorithmic systems—as well as potential mitigation efforts. Several common problems are surveyed below, which the task force urges readers to consider for each system in the context of any expected benefits and the legacy decision-making process that the system might supplant.[60]

ERROR RATES

There are many examples of algorithmic systems with high error rates, with varying and diverse causes.

Facial-recognition systems, in particular, have been plagued by errors. Some jurisdictions across the country are embracing this technology—which has drawn the ire of many researchers and criminal-justice reformers—while others are seeking to curtail it.[61]

In a 2018 study of automated facial-recognition algorithms and associated data, Joy Buloamwini of MIT Media Lab and Timnit Gebru, then of Microsoft Research, found startling error rates along racial and gender lines. Their research found error rates between 20.8% and 34.7% for "darker-skinned females"—"the most misclassified group"—in an evaluation of three "commercial gender classification systems."[62] Additionally, they found that the systems performed better on lighter faces than darker faces (11.8%–19.2% difference in error rate) and on male faces than female faces (8.1%–20.6% difference in error rate).[63] Critically, the report observed that the relevant datasets serving these systems were disproportionately comprised of "lighter-skinned subjects."[64]

In congressional testimony, Buloamwini highlighted how these broader error patterns can lead to a devastating case of misidentification in real life. In spring 2019, Sri Lankan authorities, using facial-recognition technology, falsely identified Amara K. Majeed (an activist and then a Brown University senior) as a terrorist suspect implicated in the Sri Lankan Easter bombings, leading to death threats and police scrutiny of her family in Sri Lanka.[65] In a startling episode in 2018, the ACLU ran photos of members of Congress through Amazon's Rekognition facial-recognition offering and found flawed results: 28 members incorrectly matched with arrestees' mugshot photos.[66] Moreover, "Of the false matches, 39 percent were people of color, even though people of color make up only 20 percent of lawmakers in Congress."[67] In June 2020, as racial-justice protests and calls for greater police accountability swept the country, Amazon announced a "one-year moratorium on police use of" Rekognition.[68] That suspension has continued, and other tech companies have similarly stepped away from offering facial-recognition technologies to law enforcement.[69]

And in Detroit, in the wake of a *New York Times* account of a man wrongfully arrested in January 2020 after an incorrect algorithmic identification, the police chief there acknowledged that the facial-recognition system in question had a remarkably high error rate, saying: "If we would use the software only [to identify subjects], we would not solve the case 95-97 percent of the time. … If we were just to use the technology by itself, to identify someone, I would say 96 percent of the time it would misidentify."[70]

Meredith Whitaker of New York University's AI Now Institute, who also presented to the task force, testified before the U.S. House Committee on Oversight and Reform that "[f]acial recognition reflects and amplifies historical and present-day discrimination."[71]

Facial-recognition systems are not the only systems that can suffer from inaccurate results. Analysis by *ProPublica* found that Broward County, Florida, relied on an algorithmic system that falsely labeled Black criminal defendants as nearly twice as likely to re-offend as White criminal defendants.[72] That system (called "COMPAS") reportedly relied on data that included questions about parents' incarceration history—a racially fraught piece of data with obvious implications for bias and fairness, particularly in areas with legacies of racially biased criminal-justice systems.[73] A subsequent study by Northpointe Inc. (the company behind COMPAS) countered and offered evidence refuting "the claim that the COMPAS risk scales were biased against black defendants in a sample of pretrial defendants in Broward County, Florida."[74]

BIASED DATA

A common question from the public posed to the task force was, "What information informs the tool?" There may be no more important question to ask in examining the potential for bias in algorithmic system–based decision-making. Algorithmic systems—like human processes to make decisions in government—can, and often do, suffer from bias embedded in the data on which such systems rely.[75]

NON-REPRESENTATIVE DATA

Non-representative data will undermine any algorithmic system. As Alice Feng and Shuyan Wu observed in the *Parametric Press,* "A non-representative sample where some groups are over- or under-represented inevitably introduces bias in the statistical analysis."[76]

This excerpt from a New York Times report tells the story of a man wrongfully arrested due to a facial algorithm's error.

On a Thursday afternoon in January, Robert Julian-Borchak Williams was in his office at an automotive supply company when he got a call from the Detroit Police Department telling him to come to the station to be arrested. He thought at first that it was a prank.

An hour later, when he pulled into his driveway in a quiet subdivision in Farmington Hills, Mich., a police car pulled up behind, blocking him in. Two officers got out and handcuffed Mr. Williams on his front lawn, in front of his wife and two young daughters, who were distraught. The police wouldn't say why he was being arrested, only showing him a piece of paper with his photo and the words "felony warrant" and "larceny."

His wife, Melissa, asked where he was being taken. "Google it," she recalls an officer replying.

The police drove Mr. Williams to a detention center. He had his mug shot, fingerprints and DNA taken, and was held overnight. Around noon on Friday, two detectives took him to an interrogation room and placed three pieces of paper on the table, face down.

"When's the last time you went to a Shinola store?" one of the detectives asked, in Mr. Williams's recollection. Shinola is an upscale boutique that sells watches, bicycles and leather goods in the trendy Midtown neighborhood of Detroit. Mr. Williams said he and his wife had checked it out when the store first opened in 2014.

The detective turned over the first piece of paper. It was a still image from a surveillance video, showing a heavyset man, dressed in black and wearing a red St. Louis Cardinals cap, standing in front of a watch display. Five timepieces, worth $3,800, were shoplifted.

"Is this you?" asked the detective.

The second piece of paper was a close-up. The photo was blurry, but it was clearly not Mr. Williams. He picked up the image and held it next to his face.

"No, this is not me," Mr. Williams said. "You think all black men look alike?"

Mr. Williams knew that he had not committed the crime in question. What he could not have known, as he sat in the interrogation room, is that his case may be the first known account of an American being wrongfully arrested based on a flawed match from a facial recognition algorithm, according to experts on technology and the law.

Source: Kashmir Hill, "Wrongfully Accused by an Algorithm," New York Times (June 24, 2020), https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html.

Such errors are introduced where the difference between the sample set and the broader population is a result of the sample itself; in other words, a sample error exists where a sample is not representative. Feng and Wu illustrate a sampling error by demonstrating how a sample would be skewed if, in estimating the average U.S. household income, the sample of 100 residents included billionaire Jeff Bezos (an overestimate). Conversely, a sample that included mostly low-income residents would lead to an underestimate. [77]

As another example, imagine an algorithmic system that relied on health data derived only from public health sources (e.g., Medicaid, Veterans Affairs) while excluding data from private health insurers. Or even if such an algorithmic system included private health insurers in its data sources but excluded sources such as free clinics or physicians who do not accept insurance, the system could be relying on non-representative data.

Non-sampling errors can also make data non-representative. Inaccurate or missing survey responses, poor data-collection practices, and the like can cause this type of error.

REPRESENTATIVE BUT STILL BIASED DATA

Even where sampled data are representative of broader data, bias can nonetheless exist. This can commonly occur where historical biases are "baked into" data by their very nature. In one high-profile example, such errors caused a widely used healthcare algorithmic system to suffer from bias.

In that case, researchers uncovered substantial bias in an algorithmic system that health systems and insurers nationwide (covering approximately 200 million people) rely on "to target patients for 'high-risk care management' programs."[78] Because the algorithmic system predicts *healthcare costs,* rather than *illness,* "At a given risk score, Black patients are considerably sicker than White patients."[79] In other words, the algorithmic system was holding Black patients to a higher sickness threshold than White patients in its scoring of patients to determine need for high-risk care management. The authors of the study attributed the outcome to biased data: "Unequal access to care means that we spend less money caring for Black patients than for White patients," thus artificially deflating the predicted healthcare costs for Black patients and weakening the factor's value as a proxy for predicting illness,[80] while perhaps modeling and problem formulation decisions were also responsible for the outcome. Consequently, the data here may have been representative—there are, in fact, profound disparities in healthcare costs along racial lines—but the data (coupled, perhaps, with modeling and problem formulation decisions) nonetheless introduced bias because of the stark racial differences in healthcare access that drive healthcare costs for patients.

Historical arrest data may also be representative but nonetheless racially biased due to racial disparities in policing and enforcement practices. Representative data that capture higher rates of arrest for Black residents would nonetheless be biased when the effects of historically racially biased policing and prosecution are considered. Prominent data scientists and researchers caution that criminal-history data, in general, "is neither a reliable nor a neutral measure of underlying criminal activity" and when used for risk-assessment tools create distorted, inaccurate, and racially biased results.[81] Criminal history data reflect patterns of policing and prosecution in addition to, or rather than, actual criminal activity.

For example, a report from the Center on Race and Social Problems at Pitt's School of Social Work found that "Blacks and Whites have comparable drug use rates but Blacks have much higher arrest rates."[82] The report observed that both racial groups "are equally likely to sell and use drugs," yet "Black youths were arrested for drug violations at a rate nearly five times that of White youths in the city of Pittsburgh and Pittsburgh [Metropolitan Statistical Area], and at more than nine times in Allegheny County."[83] There was a similar disparity for adults: "Black adults were arrested at four times the rate of White adults for drug violations in the city of Pittsburgh, five times in Allegheny County, seven times in the Pittsburgh [Metropolitan Statistical Area], and three times in the nation."[84] Such a disconnect between criminal activity and arrests for that activity yields bias in arrest data—making any representative sample of that data nonetheless biased, thereby infecting any system relying on that data.

OTHER SYSTEM ISSUES

Algorithmic systems can also churn out biased outcomes if machine-learning algorithms themselves produce decisions even more slanted than training data. Incentives to reinforce majority-group observations and feedback loops both can facilitate this problem.[86]

With respect to the former, "[M]achine learning algorithms are incentivized to put more learning weight on the majority group, thus disproportionately predicting observations to belong to that majority group."[87] Feedback loops can also lead to amplified bias in algorithmic system predictions, exacerbating existing bias.[88] As one dictionary describes it, a feedback loop is "the path by which some of the output of a circuit, system, or device is returned to the input."[89] In essence, the input to a system ultimately affects the system's output, which in turn then affects the inputs, and so on.

///////////////

A NOTE ON PRIVACY

Legal scholars are often challenged by the law's inability to keep up with the ever-evolving nature of technology. So too for privacy advocates.

Algorithmic systems can heighten concerns about government surveillance, especially to the extent such systems use so-called "Smart Cities" data—that is, data mined from a network of sensors to surveil through tagging, touching, tracking, and the like. Such practices raise the question of whether these digital contacts collide with Fourth Amendment doctrine regarding government "searches."[85] Law Professor Andrew Guthrie Ferguson, for example, suggests that such acquisition of personal data triggers scrutiny under the Fourth Amendment and might violate the Constitution.

Although resolving such constitutional issues is beyond the scope of this report, the task force is nonetheless sensitive to the privacy implications of the data collection often underlying public algorithmic systems and encourages embedding privacy protections alongside data-collection rules.

Feedback loops can perpetuate harms in current systems and structures, such as the criminal-justice context: "Deploying predictive policing systems in jurisdictions with extensive histories of unlawful police practices presents elevated risks that dirty data will lead to flawed or unlawful predictions, which in turn risk perpetuating additional harm via feedback loops throughout the criminal justice system."[90]

LACK OF TRANSPARENCY

Beyond the substantive problems that can plague algorithmic systems, a lack of transparency can undermine public faith and confidence in governmental use of algorithmic systems. The popular notion of an algorithmic "black box" has some truth to it, and people are loath to trust something that feels concealed from public scrutiny or where accountable elected officials have no role in oversight.

A lack of transparency can result from a lack of public oversight of—or insight into—algorithmic systems and underlying data. Many private algorithmic-system vendors on which governments rely do not make available their systems or training data for public review. Similarly, such companies typically do not explain how systems reach decisions in an effort to keep their algorithms from being reverse-engineered. Both the public and elected officials are blocked from effective oversight of systems under those conditions.

///////////////

Consider this example from Florida, where the Pasco County Sheriff's Office "keeps a secret list of kids it thinks could 'fall into a life of crime' based on factors like whether they've been abused or gotten a D or an F in school."[91] The data are drawn from school district records and state Department of Children and Families information. Neither the children nor their parents are informed when children are added on the list or how the system weighs or treats the data, which is especially problematic given disparities in discipline at the school district: Black students and students with disabilities in Pasco County are twice as likely to be suspended or referred to law enforcement.

Source: Neil Bedi & Kathleen McGrory, Pasco's sheriff uses grades and abuse histories to label schoolchildren potential criminals, Tampa Bay Times, Nov. 19, 2020, https://projects.tampabay.com/projects/2020/investigations/police-pasco-sheriff-targeted/school-data.

So too is such oversight likely thwarted where review of the substantive judgments inherent in a system are shielded. Neither the system nor the data[92] when kept obscured from the public can provide adequate transparency because algorithmic tools make a series of determinations "about what data to use, include or exclude, how to weight the data, and what information to emphasize or deemphasize."[93] According to law professors Robert Brauneis and Ellen Goodman (a task force member), "[T]here are three principal impediments to making government use of big data prediction transparent: (1) the absence of appropriate record generation practices around algorithmic processes; (2) insufficient government insistence on appropriate disclosure practices; and (3) the assertion of trade secrecy or other confidential privileges by government contractors."[94]

Sufficient transparency allows the public to ensure that a system is making tradeoffs consistent with public policy. A common tradeoff is balancing the risk of false positives and false negatives. A programmer may choose to weigh those in a manner different than policymakers or the public might prefer. An example from Philadelphia highlights this tension: "Philadelphia's Adult Probation and Parole Department's risk prediction algorithm for violent recidivism among probationers … predicts the likelihood of a probationer committing a violent crime within two years of release, and classifies the population as high, medium, and low risk. The algorithm was constructed by treating historical false negatives as 2.6 times more costly than false positives."[95] After examination, officials changed that rate to better align with department goals to ensure that the system was not overclassifying probationers as high risk.[96]

LACK OF PUBLIC OWNERSHIP

When a public agency procures a system from an outside vendor—rather than develops a system internally or with assistance from an outside vendor or expert—the agency typically does not own the system, or sometimes even the data. Consequently, the governmental entity ends up with less control and authority over the system, including directing changes and revisions to the system, than it would have over a system that it owns.

DESIGNER DEMOGRAPHICS AND DESIGN CHOICES

The community of those who prepare, develop, and implement algorithmic systems is disproportionately filled with those with advanced degrees (typically men) who often do not look like or share the lived experience of the people affected by the systems they design.[97] Such demographic divides can undermine developers' understanding or ability to appreciate the problems or lived experiences of those subject to algorithmic systems.

OVERWEIGHTING ALGORITHMIC SYSTEM OUTCOMES

When algorithmic systems' outcomes are merely suggestive in nature—as opposed to mandatory—a human component of decision-making is still critical to the process or function at issue. Yet research suggests that humans place inconsistent weight on algorithmic systems' outcomes, with those from less privileged backgrounds and wielding less power exerting less independence in disregarding algorithmic systems' outcomes.[98] Relatedly, when humans place too much confidence in algorithmic systems' outcomes, algorithmic systems can exert undue influence on decision-making processes that are supposed to incorporate more human judgment,[99] and humans might rely on algorithmic suggestions more heavily when the suggestions reinforce their own inclinations.[100]

INADEQUATE PUBLIC INPUT AND DIRECTION

The absence of public input and direction regarding the development and embrace of algorithmic systems (a corollary to the transparency issues discussed above) is a common failure of governmental introduction of algorithmic systems. Task force community feedback and scholarship identified this problem, which can undermine public acceptance of and trust in algorithmic systems, regardless of how effective a system might be.

The task force concurs with this observation from the Centre for Data Ethics and Innovation:

"A range of transparency measures already exist around current public-sector decision-making processes. There is a window of opportunity to ensure that we get transparency right for algorithmic decision-making as adoption starts to increase."[101]

In Boston, for example, the public school system looked to an algorithm to improve bus schedules: more efficiency in routes, reduced excess bus travel, better accommodations for disabled students, optimized school start times, and equity considerations. Despite a community-engagement process and close work with the city, this "marvel" of an algorithm nonetheless met "strong and swift" "pushback" from the public.[102] As task force member Ellen Goodman wrote in describing the incident, "The consultations had not effectively described how the start time shifts in particular would change students' schedules and they had not given families opportunities to play with the model," and the protesting public blamed the algorithm—which could have been tuned to optimize for other objectives—for failures of political process.[103] This episode highlights an important consideration for policymakers: Even when an algorithmic system might improve outcomes on certain metrics, the public's distrust or rejection of the system can imperil it.[104]

This section surveyed some of the problems that algorithmic systems might carry into government decision-making. Some of those problems might already exist (or even be worse) in legacy decision-making. Consequently, the task force urges stakeholders to take a proper accounting of the expected costs and benefits of an algorithmic system—weighed against the process it would supplant—when considering such a system.

IV. Task Force Recommendations

Our regional governments' embrace of algorithmic systems—and the growth of those systems—is happening against a backdrop of disparities in transparency and oversight across different agencies of regional governments. Moreover, without caution, these systems have the potential to perpetuate or cause harm and inequity, especially for members of marginalized communities and groups, or to weaken trust in government. In some careful applications, algorithmic systems can offer the promise of improving upon legacy decision-making processes. It is clear to the task force that our region is at a moment ripe for examination and action to chart a responsible and consistent path forward. That path ought to be one that not only confronts the biases and harms that might flow from some systems but also works to build trust in government.

There is no meaningful, concrete oversight effort regarding algorithmic systems in place in this country which the task force could simply look to adapt to our region. To be sure, there are other models and proposals from which to learn to varying degrees,[105] and the task force sought to learn from experience and experts both within and beyond our region.

The practices below, if adopted, will not fundamentally restructure systems of injustice and oppression where they exist, but they are designed to make more than just incremental improvements in governmental processes. Our hope is that these recommendations will push agencies to foster public participation, bolster meaningful scrutiny of agencies' algorithmic systems, and set conditions for improved government decision-making.

Our hope is that these recommendations will push agencies to foster public participation, bolster meaningful scrutiny of agencies' algorithmic systems, and set conditions for improved government decision-making.

RECOMMENDATIONS FOR LOCAL GOVERNMENTS

RECOMMENDATION 1

1 Encourage meaningful public participation, commensurate with the risk level of the potential system.

The task force believes that not all algorithmic systems are alike and that each presents different levels of risk. Moreover, such risks exist in general in government decision-making, irrespective of the deployment of an algorithmic system in that decision-making process. Some systems (and the decisions they shape) implicate liberty, others affect the allocation of resources, and still others might simply influence mundane public functions such as management of traffic patterns. These varying risk levels, in the view of the task force, ought to inform the depth of government efforts to bring the public in to participate in the lifecycle of a system (and the decision-making process it will inform).

Consequently, government leaders should understand the relative risk of any given system as an initial matter. To assist in that assessment, the task force found the framework proposed by Gretchen Greene of the AI and Governance Assembly to be useful. Under the framework, levels of risk are defined as "the likelihood of causing serious harm through discrimination, inaccuracy, unfairness or lack of explanation."[106] The table below from the framework presents some common algorithmic systems and their corresponding risk levels:

LOW RISK			HIGH RISK
Infrastructure maintenance and repair	Disaster management	Regulatory inspections	Criminal justice Entitlement decisions Firing decisions Child protection services
Potholes (Boston)	Wildfire spread prediction (U.S. Forest Service) Landslide detection (NASA–rescue for isolated villages)	Restaurants (Chicago, Boston) Building fire code (Atlanta, New Orleans)	Predictive policing Court sentencing

Source: Gretchen Greene, Potholes, Rats and Criminals: A Framework for AI Ethical Risk[107]

The framework also provides a series of questions designed to provide insights into the "kind and seriousness of the possible harm" that the system might cause.[107] In the table below, answers are provided for common examples of algorithmic systems, demonstrating how such examples are assessed at different levels of risk:

	LOW RISK			HIGH RISK
	Pothole detection phone sensor app	**Landslide detection (NASA–rescue for isolated villages)**	**Restaurants health department inspection scheduling**	**Criminal court risk score used in sentencing**
Use of citizen data	Minimal or none	Minimal or none	Low	Medium to high
Citizen data use opt in/ opt out	Opt in	No opt out	No opt out	No opt out and possible self incrimination
Use of class data or proxies where class has special antidiscrimination legal protections	None	None	Minimal	Medium (no direct race input but studies have shown results vary by race)
Seriousness of worst possible citizen harm, e.g. loss of life, liberty, or property	Minimal or noon Secondary/indirect cause	High (harm = not found for rescue but compare to status quo) Secondary/indirect cause	Low Secondary cause	High Direct government taking of liberty or property
Barriers to citizen challenges to the data/ algorithm as discriminatory, inaccurate, unfair or lacking explanation	Low (if alternate method exists to report potholes)	High	High	High

Source: Gretchen Greene, Potholes, Rats and Criminals: A Framework for AI Ethical Risk[107]

The task force does not believe that these questions need to be rigidly followed; however, this framework presents a workable, flexible approach that could help agencies understand algorithmic system risk. Armed with that risk assessment, agencies could thus undertake a public process tailored to the system's expected risk. The task force encourages agencies to include in risk assessments public consultation about risk level.

As described below in Recommendation 2, agencies should publish baseline information about the proposed system: what the system is, its purposes, the data on which it relies, its intended outcomes, and how it supplants or replaces existing processes, as well as likely or potential social, racial, and economic harms and privacy effects to be mitigated. That disclosure can be a jumping off point for agencies to invite the public to participate, consistent with the system's risk level. Such engagement could include public forums, social media campaigns, diverse advertising, and other opportunities for members of the public to comment on the system. *For the highest-risk systems, agencies should consider the whole range of options to reach members of the public. An emphasis on communities and organizations representative of those most likely to be affected by the system is essential.* Outreach should invite the public in to discuss whether the potential system's goals and purposes are embraced by the community, and outreach should continue after system deployment (especially regarding any substantive changes or revisions to the system).

> "Approach the community at the onset of a problem where an algorithm might be a solution."
>
> —Community member attending public task force meeting

There is a key distinction to be drawn between public-comment opportunities—which would be insufficient for higher-risk systems—and more robust public participation, which the task force encourages. As a resource, agencies might look to the work of researchers from New Zealand and the United States (including task force member Alexandra Chouldechova) who conducted a study to learn through workshops "about the concerns of affected communities in the context of child welfare services."[108] Their insights and findings would be invaluable to shaping effective participatory design approaches, as would consulting New Zealand's national guidelines for obtaining social license for data use, which were informed by public participation.[109]

For the lowest-risk systems, the task force believes that it will usually be sufficient for agencies to offer straightforward comment avenues along with public disclosure and input opportunities.

RECOMMENDATION 2

2 Involve the public in algorithmic system development plans, from the earliest stages through any later substantive changes to the system.

Algorithms can be mystifying to people, leading to distrust of, or overreliance on, a tool or system that is not easily understood or assessed—often irrespective of the system's performance. Although transparency alone cannot unlock the complexities of machine learning and the like, the value of openness could serve both to promote public participation and knowledge and to allow greater scrutiny of algorithms. This is especially important in the context of public algorithmic systems—the focus

> "An algorithm would be useful to identify the root problems and to allocate the right services, training, and resources needed to mitigate the problem."
>
> —Community member attending public task force meeting

of the task force—where governments are relying on these systems to perform or assist public functions that impact residents' lives.

Hetan Shah, for example, argues that the opportunity for positive social impacts from algorithmic systems requires a "licence to operate" from the public.[110] This "licence to operate" (or social license) includes a role for transparency.[111] There is further scholarship focused on how best to foster meaningful debate about whether an algorithmic application is acceptable (and how best to participate in the development of such systems), grounded in a recognition that such public involvement is required to obtain social license.[112] The example discussed above from Boston—where a seemingly effective school-busing algorithm was nonetheless imperiled by strong public pushback—highlights the consequences of inadequate, insufficient, or non-responsive engagement with the public: public distrust and potential abandonment of a system.

The public shared with us many concerns regarding transparency (or a lack thereof) with algorithmic systems. Even in a community session with "grasstops" leaders, there was a sense of surprise at the scope of algorithmic systems already in use and a perception of a lack of engagement with the public when those systems were introduced. One relevant public query asked, "Are the right organizations sharing information [about their algorithmic systems with the public], and what are they doing with it?"

The task force's community engagement yielded a sense from residents that government use of algorithms could be useful, but there were concerns that the systems would not be developed in a way that reflected community values and needs.

In particular, residents are concerned that algorithms have been and would continue to be deployed in primarily punitive efforts. There was support, however, for the development of algorithms that are used to identify resource gaps and allocate resources accordingly.

To improve awareness, the task force believes that government agencies should alert the public and government leadership of any agency plan to develop, procure, or deploy an algorithmic system, with an eye toward involving the public in the decision about whether to pursue the system. Such notification should include key information to be meaningful, for example:

- Identification of the system and its purposes
- Description of the data on which the system relies
- The system's intended outcomes and benefits
- How the system supplants or replaces existing processes
- Likely or potential social, racial, and economic harms and privacy effects to be mitigated
- Identification of any non-governmental funders of the system

> "It seems like a lot of algorithms are used from a deficit model of thinking. 'Here are the problems.' But where are the models for seeing our strengths? Where are the conditions under which you can predict health outcomes for children, whether it's housing or air quality? We could use that to site affordable housing and prioritize affordable housing. Where are the strengths, and where should we be leveraging these strengths?"
>
> — Community member attending public task force meeting

Beyond mere notification, proactive engagement of the public, as reinforced and discussed above, is important from the outset by any agency considering use of an algorithmic system. The public should be involved in the deliberation over whether and how to pursue a system—including the policy goals at issue—as well as later efforts to shape, refine, and improve the system.[113] Indeed, as AI Now Institute has observed, such efforts "[r]espect the public's right to know which systems impact their lives.[114] Moreover, as described below in Recommendation 5, the task force believes that local governments ought to publish information about algorithmic systems on a public website.

RECOMMENDATION 3

3 Utilize third-party reviews when a system might be higher risk.

The task force believes that agencies should subject higher-risk algorithmic systems to outside, independent review.[115] Such reviews—which should be contemplated by agency leadership well before system deployment—can offer two fundamental benefits.

First, a review can be useful at all stages, including identifying shortcomings, weaknesses, errors, biases, and the like in the critical stage *before* full implementation. This can arm government leaders and researchers with information—gleaned from those outside the team behind the system—which they can use to improve a system or mitigate issues.

Second, a review can provide not only an added layer of visibility into a system but also a basis on which public trust might be earned. A system backed up by a fulsome third-party review might be more likely to engender public trust.

The task force believes our region is fortunate that county DHS has already established a track record of soliciting and publishing these types of reviews.[116] Other agencies ought to follow the lead of DHS in incorporating reviews into the development of algorithmic systems. Thus, we encourage agencies to obtain and publish for each algorithmic system:

- Ethical review
- Process review[117]
- Impact evaluation

Earlier in this report, we detailed the DHS's AFST and made more specific reference to those reviews, which provided important independent scrutiny of the AFST and DHS in a publicly accessible manner. Agencies might also consider obtaining other reviews as appropriate (including, for example, a data-science review to examine data practices, privacy, and security, identifying ways malicious actors could subvert the system, and more). Moreover, there can be tremendous value in ensuring that a system's deployment is set up in a way that facilitates ongoing and meaningful impact evaluation post-deployment.

RECOMMENDATION 4

4 Integrate algorithmic review into procurement processes.

Local governments in the Pittsburgh region already have established processes to manage procurement and contracting for outside services. The task force believes that governments could leverage those existing frameworks by adding a review to assess whether any planned procurement might include an algorithmic system.

Similarly, that review could also assess whether public data will be used in a way suggesting that an algorithmic system is at play.

This review—built into existing processes with which government officials have familiarity—would provide an internal check to ensure that algorithmic systems are identified for government leadership and the public. When such a review flagged a system slated for procurement, if public notification had not already occurred (as outlined above in Recommendations 1 and 2), then government officials would have the information needed to make that notification, both internally and externally.

For example, the City of Pittsburgh requires that every procurement contract goes before city council after an agency has selected a vendor. Consequently, city officials should conduct their reviews *before* contracts are presented to council (i.e., before legislation is presented to council for the contract's budget allocation), thereby enabling members and the public to review an agency's planned acquisition more effectively. Moreover, the city's Department of Innovation and Performance, which handles software acquisition (including algorithmic systems), already has language in its Data Governance Operational Charter that seeks to protect civil liberties and avoid bias in algorithms: "We recognize that in achieving our mission, it is critical that we remain committed to ensuring that our usage of data does not in any way infringe upon the privacy or civil liberties of citizens, and that we maintain accountable and bias-free utilization of computational algorithms."[118]

Local governments should consider updating existing checklists and workflows and training those involved in procurement to account for a review of potential algorithmic systems and/or use of public data suggestive of an algorithmic system. Familiarity with the definition of a *public algorithmic system a*nd likely applications will be key.

RECOMMENDATION 5

5 Require agencies to publish information about algorithmic systems on a public website.

The task force believes that the public would be well served by local governments maintaining a public-facing registry website where such information about algorithmic systems would be available.[119] Such a website should include, in plain language, the system-specific information included in Recommendations 1 and 2 above.

A subset of the task force is developing a website prototype—to be informed by public participatory design—that could be the basis for city and county websites. At a minimum, any relevant website should list the information described in Recommendation 1 above for each algorithmic system and provide agency points of contact. As was similarly recommended by New York City's Automated Decision Systems Task Force, local leadership should also consider making the information available in printed form at Carnegie Library branches and other culturally relevant sites, ensuring broad access even where internet access might be limited.

RECOMMENDATION 6

6 Avoid facial recognition and related systems.

Section II above documented common errors—often with profound racial and gender disparities—associated with biometric algorithmic systems such as facial recognition. The potential for harm is heightened because governments typically use facial and related systems (e.g., affect recognition) in high-risk applications that, even if accuracy issues eventually improve, could result in invasive surveillance that would undermine privacy. Consequently, based on the current state of these technologies, the task force recommends that governments avoid such systems for the foreseeable future.

This is of particular relevance in Allegheny County, where there is a network of more than a thousand surveillance cameras (which reportedly pose a national security threat due to their manufacturing origins in China) and the possibility of implementation of facial-recognition technology.[120] Additionally, in connection with the 2020 racial-justice demonstrations in Pittsburgh, the Bureau of Police used a state facial-recognition system (known as "JNET") to match a social-media image to one in the database in order to identify a person they charged with crimes.[121] The bureau apparently disregarded an agency policy that it "does not use facial recognition software or programs,"[122] highlighting the urgent need for more robust oversight of facial-recognition capabilities. In the aftermath of the revelations about the Bureau of Police's reliance on JNET, Pittsburgh City Council approved a bill in September 2020 to modestly regulate facial-recognition technology, although the bill included an exception permitting use of JNET.[123] Yet in light of the bureau's own identification of officer disciplinary action for JNET-related misconduct,[124] that JNET exception seems troubling.

RECOMMENDATION 7

7 Evaluate effectiveness of recommendations.

The task force appreciates that the experience of local governments with these recommendations ought to shape the future. There will inevitably be successes and failures. We urge stakeholders to learn from those lessons in revisiting how best to manage and oversee public algorithmic systems in our region.

Perhaps these recommendations will fall short, and there will be public support for more robust and mandatory regulation. Or perhaps these recommendations will hit the mark in terms of fostering innovation, ensuring that algorithmic systems do not perpetuate existing biases, and encouraging agencies to develop algorithmic systems through the lenses of economic, racial, and social justice. Regardless, the task force believes that our region should leverage the knowledge that comes from implementation of these recommendations after a year of implementation by evaluating how they might be improved in the future (perhaps through a deliberative body within government or, like the task force, outside it).

V. Additional Best Practices

The task force's recommendations described above, if adopted, could provide a balanced approach to oversight of government use of algorithms and government innovation. They focus largely on instilling public participation and transparency, both necessary preconditions to just algorithms. The task force also offers the below suggestions for best practices for municipal governments more broadly. Some are inherent in the proposed recommendations above, but the task force thought it worthwhile to articulate them specifically. Others do not fit directly into the discussion above but are well worth considering.

AGENCIES

ASK: DO WE NEED AN ALGORITHM?

It is easy to think an algorithm is your solution—and vendors are often happy to sell an agency one. We anticipate a pattern in the future where algorithms are seen as easy answers to hard problems but without necessarily improving existing efforts. As with any significant government expense, hard questions should be asked about whether an algorithmic system will improve residents' lives more than an alternative effort and/or an existing process. And, as discussed elsewhere in this report, algorithms are not solutions to systemic problems.

BE CAREFUL WHAT YOU BUY—AND AVOID MISSION CREEP.

Increasingly, vendors are selling software that includes algorithmic systems an agency might not intend to purchase (or that relies on data gathered for one purpose but then deployed for an altogether different one). But once an agency has a capability, it is hard not to use it. We note, of course, that if the task force's recommendations were adopted, a new or expanded use of a system, even in previously procured software, should be disclosed.

Likewise, once an agency has an algorithm, however procured, potential new use cases will be identified. These also should proceed through disclosure and related processes, as appropriate.

PROACTIVELY ADDRESS VENDOR ISSUES.

The recommendations are designed to ensure significant transparency, incentivizing agencies to develop explainable public records around algorithm creation and use. However, when a public agency procures a system from an outside vendor—rather than developing the system internally or with assistance from outside vendors or experts—the public agency typically does not own the system. Consequently, government ends up with less control and authority over the system, including the power to direct changes and revisions, than it would have over a system that it owns.

As task force members Professor Ellen Goodman and Professor Robert Brauneis have written, this can lead not only to "opacity, public disempowerment, and loss of accountability" but also to algorithms that "may enact policy judgments that diverge from the preferences of the electorate or its elected representatives."[125]

As algorithms become more common in municipal governance, we anticipate that they will often be procured from vendors. At a minimum, when working with an outside vendor, agencies should (1) require that the system is auditable; (2) maintain ownership and access to any government data; (3) require that appropriate records are generated; and (4) contractually limit claims of trade secrecy.

Agencies should also carefully consider hiring or assigning personnel with responsibility for algorithmic system issues. Such responsibility should include not only facilitating public participation, but also spearheading agency efforts in guiding the development and/or procurement of algorithmic systems.

TRAIN YOUR PEOPLE.

Any algorithmic system should be interpretable to the public—as well as to the employees who use it. Agencies must ensure that employees understand any limitations regarding what a system was designed for. For example, we have seen in other jurisdictions that systems designed only for pre-trial detention decisions are sometimes then used for sentencing, a use the developer of those systems never intended and did not build them for. An agency and its employees should know what a system does—and does not do—and how reliable it really is. Employees using such systems often have little sense of accuracy rates of systems, essential information if one is relying on it to make a decision. This will also help limit deskilling of an agency's workforce.

Relatedly, agencies should be aware of the risks of automation bias and train their employees accordingly. When algorithmic system outcomes are suggestive in nature—as opposed to mandatory—a human component of decision-making is still critical to the process or function at issue. When humans place too much confidence in algorithmic system outcomes, such systems can exert undue influence on decision-making processes that are supposed to incorporate more human judgment.

Agencies relying on algorithms also need employees who are capable of interrogating an algorithm: e.g., scrutinizing its modeling and development, data sets, outputs, processes, and more. Relying on a developer to have done so is irresponsible at best, both in development and deployment stages.

ENSURE ETHICAL DATA COLLECTION, MAINTENANCE, AND GOVERNANCE.

Municipal governments are collecting and will only continue to expand collecting residents' data. Data are necessary for good policymaking. In particular, good (and often unbelievable volumes of) data are necessary for good algorithms. Yet, resident data raise numerous ethical and safety issues—from consent to quality control to security.

We urge stakeholders to emphasize ethical data governance, including holding sensitive data for only as long as necessary and implementing robust security safeguards.

PUBLIC

MITIGATING ALGORITHMIC BIAS WILL NOT FIX SYSTEMIC ISSUES.

Algorithmic bias is a symptom of much larger systemic issues. Mitigating it is important but no silver bullet. This task force came together because we believe it is a critical and pressing need to provide more oversight of algorithmic use in local government—something that can perpetuate and even further lock in existing injustices. However, fixing algorithms will not fix racism or poverty or injustice. When we think about algorithmic oversight as communities, as advocates, as individuals, we should approach it with this limitation in mind.

ALGORITHMS ARE NOT PERFECT, BUT NEITHER ARE HUMANS.

We should not expect perfection from our government algorithms. But we should expect that agencies are able to demonstrate that algorithmic systems produce equal or better outcomes than human processes, and there must be a way for the public to interrogate and challenge such systems.

PARTICIPATE IN THE PROCESS.

The task force hopes that our recommendations will be adopted. If so, success will depend on the depth of public participation. You do not need to be a technical expert to participate. Public deliberation and engagement are crucial in determining whether an algorithm is an appropriate tool in a particular context and whether appropriate conditions for its use have been met. You are an expert on the needs of you, your family, and your community.

OTHERS

We have intentionally not directed recommendations to developers of algorithmic systems. The responsibility of these systems in government use rests with government. However, we note that there is a large and growing body of literature and efforts around algorithmic fairness, and we encourage developers to engage with this literature and other resources thoroughly and responsibly.

However, because we in the Pittsburgh region are fortunate to have an extraordinary research community with the skills to develop algorithmic tools, it bears mentioning that we encourage that both researchers and those who might fund their work identify whether algorithms are wanted and needed both by relevant communities and agencies before embarking on efforts. Acceptance of algorithmic systems should begin with assessments of needs and capabilities and how algorithmic tools might improve governmental decision-making. As one member of the public asked at a community event, “What was the problem someone was trying to fix through developing [this] algorithm?”

Acceptance of algorithmic systems should begin with assessments of needs and capabilities and how algorithmic tools might improve governmental decision-making.

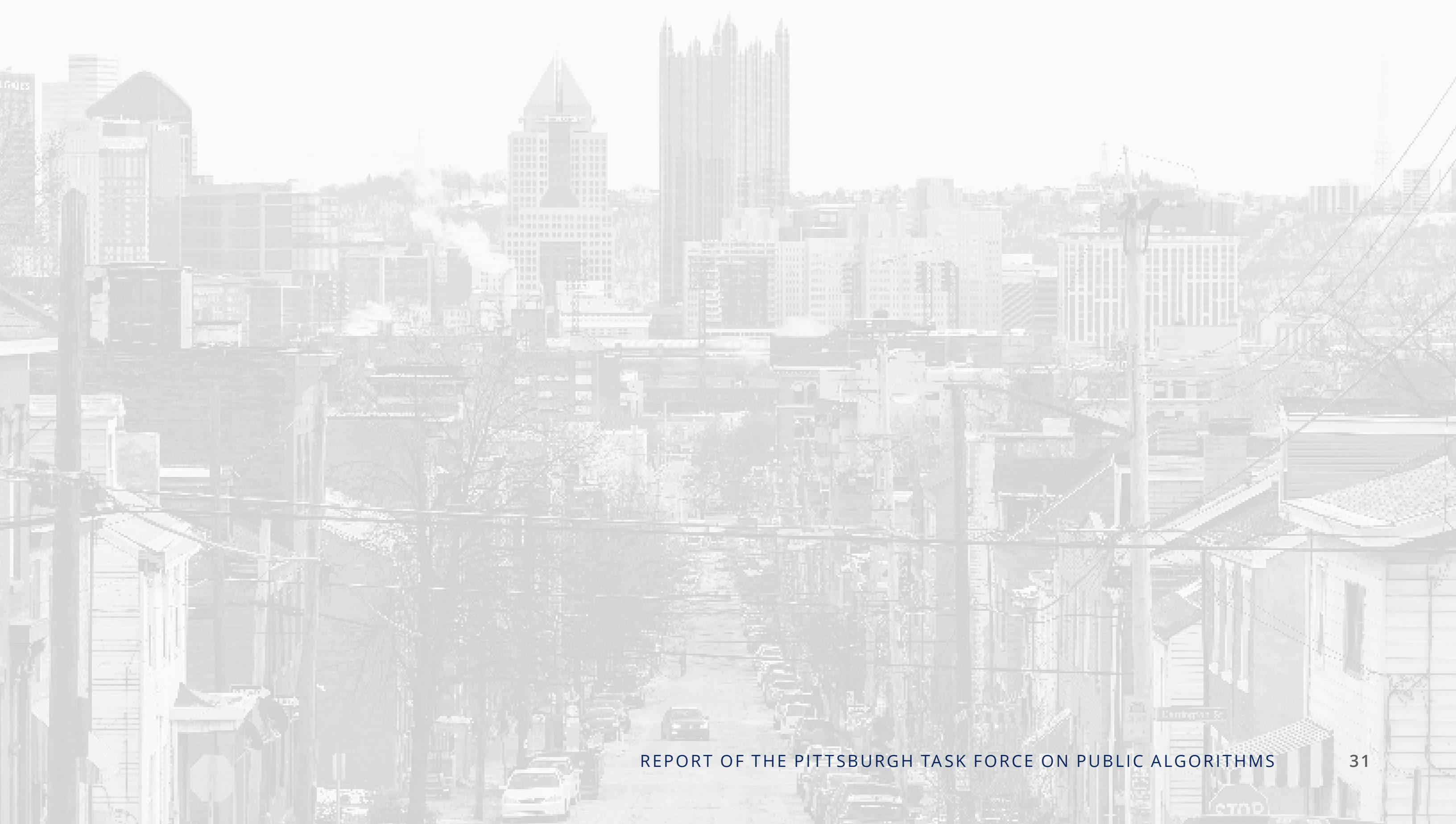
VI. Conclusion

The task force's recommendations aim to equip policymakers with actionable and concrete steps to manage and mitigate potential harms of our region's public algorithmic systems while still ensuring those systems have a chance to flourish where appropriate. Public participation is a theme throughout our recommendations, inspired by our belief that government ought to be responsive to residents.

The task force cannot advise how to resolve every conflict that might arise when an agency is deciding whether to turn to an algorithmic solution. Rather, we believe that those decisions should be made in public, with informed and open debate. This should involve people considering potential harms and their associated mitigation efforts against expected improvements relative to legacy decision-making systems. Our recommendations aim to facilitate those debates about the appropriateness of algorithmic systems in an open and public forum, one where these issues are confronted publicly, against a backdrop of fulsome public disclosures, with clear opportunities for participation and comment.

We expect and encourage deliberations on our recommendations and are confident that what we have crafted is but the beginning—and not the end—of a solution for our region. And our region's governments would be wise to work constantly to improve upon these recommendations, learning from early lessons in practice.

The task force's hope is that the Pittsburgh region will become a model for others across the country similarly confronting the proliferation of algorithms in government. The cost of sitting idly by during this transformative era in municipal government could be high, especially for the most marginalized among us.



Appendix: Community Feedback Summary

Submitted by task force member LaTrenda Sherrill, Common Cause Consultants

The governmental use of algorithms in Allegheny County is a key issue of public concern. After feedback sessions with community leaders and residents, the perspective shared by many was that although algorithms created and deployed by the government can be cost-effective time-savers, they are developed in a vacuum without community consultation. Stakeholders felt that government-produced algorithms tend to deliberately produce negative results either through purposeful bias, such as discriminatory assessment tools and data, or negligence from a lack of diverse algorithm designers. These algorithms appear to promote ways to restrict resources to communities instead of connecting the support needed to change the outcomes communicated by the algorithms. The task of making decisions about what is needed in a community should be a collaborative effort and cannot be achieved without transparency, accountability, and effective public education. This research aimed to understand public perception of municipal algorithms and the community's desired role in their use. Overall, community members posed key questions for decision-makers to consider:

- How might we create an algorithm that reflects what we value as a society?
- How might we educate our community on the purpose of an algorithm and its continued use?
- How might we include the voices of those who are most impacted by an algorithm in the process of its development?

OVERVIEW OF THE COMMUNITY-ENGAGEMENT PROCESS

As algorithms are deployed and developed throughout the Pittsburgh region, there are concerns that few efforts to ensure transparency and accountability are being deployed alongside them. To understand this fully, outreach began in the spring of 2020 to solicit community feedback. The task force originally planned four initial meetings; however, only two were possible due to Pennsylvania's stay-at-home order. Outreach continued during the summer months, which led to a number of virtual gatherings (in addition to two in-person community meetings before COVID-related shutdowns), including a virtual workshop at the University of Pittsburgh's Pitt Diversity Forum, one guest appearance on a Facebook Live show, and one virtual community meeting.

For the first meeting, more than 100 organizational leaders were invited to participate in a conversation at the Community Engagement Center in the Homewood neighborhood of Pittsburgh. For the second meeting, which also took place in Homewood, the task force attracted public participation through social media. In total, 47 community members and organizational leaders participated in the two sessions.

In addition, the task force contributed to a workshop titled "Equity and Justice in Government Algorithms," which occurred as part of the Pitt Diversity Forum. The workshop was led by Michelle McMurray, task force member and then-director of grantmaking at the Pittsburgh Foundation; Erin Dalton, then-deputy director of DHS and a member of our government advisory panel; Richard Purcell, English professor at Carnegie Mellon University; and LaTrenda Sherrill, lead consultant at Common

Cause Consultants and a task force member. The workshop was facilitated by Chris Deluzio, policy director of Pitt Cyber and a task force member. A number of local community members, university officials, and even those outside of Pennsylvania participated in the conversation, which totaled slightly more than 100 people.

To better understand the broader public's understanding of algorithms, the task force participated in the weekly Facebook Live conversation titled "What Black Pittsburgh Needs to Know about COVID-19." An initiative of 1Hood Media, the series began in March as a way to keep the Black community informed about events that directly affect them. Each week, the hosts, Jasiri X (1Hood), Dr. Jamil Bey (Urbankind Institute), and Cheryl Hall Russell (BW3 Consulting), interview a specialist in accordance with that week's theme. The episode in which the task force participated garnered more than 18,000 views on Facebook.

Finally, the task force attended Lawrenceville United's monthly community meeting. The group had discussed a number of issues related to public safety; the abundant interest from Lawrenceville residents resulted in a robust discussion. Ten people attended the virtual meeting, and about 20 people watched through Facebook Live.

During the meetings, we shared an overview of the Pittsburgh Task Force on Public Algorithms, and we asked attendees to provide a definition of "algorithm." The digital tool Menti allowed participants to respond to the prompt anonymously, and at first, most respondents originally associated algorithms with advanced math or science. Once there was a concrete understanding of what algorithms were, how they worked, and how they are used to make decisions in the real world, stakeholders had a chance to comment on a few key questions related to specific algorithms. These included algorithms that are currently in use by the local police, Allegheny County DHS, and the Allegheny County courts system. Some of the questions posed:

- **What assures you about the usage of municipal algorithms?**
- **What do you feel confident about?**
- **What about this work gives you pause?**
- **What is proising about this, or could use a bit more refinem**

After commenting on specific questions, participants were asked to think more broadly about algorithms and their interaction with the public. They were asked about how communities should be consulted in an algorithm's creation and use, as well as how algorithms should be deployed and overseen. Attendees had an opportunity to express how algorithms had played a part in their lives and how they should function within communities. Below is a summary of feedback gathered at the meetings.

THEMES FROM COMMUNITY FEEDBACK

REDEFINING PURPOSE

Stakeholders identified that algorithms are useful tools, if and when the goals are to support the community and identify resource gaps. They suggested that algorithms could be used to identify root problems (especially in marginalized communities) and to allocate services, training, and resources to strengthen community support systems. There were concerns that when algorithms are not used that way, they can do more harm than good. Algorithmic tools must be able to explicitly account for nuance and context. Community members suggested that many issues algorithms try to solve are not “black and white,” so there needs to be a system in place where algorithms are used alongside case-by-case considerations. It is essential to focus on how tools will positively affect communities going forward.

CODEVELOPING WITH COMMUNITY

Stakeholders recommended that local governments make more of an effort to share issues that may require objective tools with the public. Participants suggested that local governments approach the community at the onset of a problem or issue when it arises in order to share how an algorithm can be a possible solution to that issue. Before an algorithm is deployed, the government should hold meetings with representatives of community organizations and cultural groups, as well as individual residents. Public hearings, facilitated by the local government, should be conducted so that decision-makers can hear directly from citizens. An ethical framework to evaluate and assess the implementation of a public algorithm should be created and widely accepted. Communication with the public should be the priority.

Further comments suggested an independent board of experts to review algorithmic tools for appropriate use and training. It would also help to have monthly reports of successes, failures, and corrective measures regarding algorithms. Those who have lived experience with the systems that algorithms seek to improve should be part of the solution.

An example of a tool that would benefit from this kind of public collaboration is the Pre-Trial Risk Assessment Tool. Although the assessment in its current form helps to remove bias from subjective decisions and removes some of the power prosecutors and police have to influence bail-related decisions, the tool operates in a flawed way. Some of the items on the assessment are discriminatory of age, location of residence, employment, and history. The tool does not consider the reasoning behind certain behaviors (such as failing to appear for court), nor does it consider positive factors that work in favor of the defendant. One community member noted, “A system that predicts [something] negative based on history? You will get more negativity!” In order to avoid perpetuating a cycle that prioritizes biased data, reconsider the details of the assessment and consider other factors that contextualize the data points.

When developing algorithms, decision-makers need to ensure that the results of the algorithms will reach people who matter. Government leaders should work with their communities to determine which problems can fall under the purview of an algorithm and which ones cannot. A community member asked, “Which problems (if any) do you consider too high-stakes for algorithms?”

PROMOTING TRANSPARENCY AND ACCOUNTABILITY

The community expressed that it was vital for them to know who creates and controls the algorithms. As one community member asked: “When the algorithms go wrong, who is at fault?” Community members voiced that any algorithms in use must be repeatedly examined and reviewed. Comprehensive review needs to include an algorithm’s impact at technological and human levels; specifically, review needs to ensure that the tools are continuously serving their purpose. Also important, are certain groups more impacted than others based on the results of the algorithm? Unintended consequences of this work are possible, and stakeholders encourage leaders to take time to reflect and iterate as needed. There needs to be a system for updating algorithms, as well as determining how often they need to be changed. This system needs to be codified, as well.

The government should also examine which predictive models are actually working and which are not. Predictive policing is a concept that stakeholders suggest should not fall under the purview of algorithms. Because the history of policing has involved instances of bias and prejudice, the public is concerned about the possible framing bias that officers will gain based on algorithmic data. The community does not want algorithms to result in oversaturated policing. This type of algorithm optimizes arrests, which can be extremely harmful to vulnerable communities. The target of the algorithm needs to be redefined, and the government should rethink predictive policing as a way to prevent crime. It was expressed that if algorithms are showing that specific communities are "hot spots" for crime, the government should follow suit and allocate resources to those places—not send additional ways to "punish" the community. Many community members believe that an algorithmic tool would most beneficial if used for alternative prevention rather than policing.

In addition, community members emphasized that they would like to know how organizations are using the information generated from the algorithms. The Allegheny County Family Screening Tool, which purports to improve child safety through call screening, is an example of where transparency is needed from the government. Community members felt that government leaders should clarify what specific problem the tool solves and not only ensure the public that the data collected are accurate, but also that it will support the well-being of families and children.

Stakeholders expressed that there is a need to redefine and broaden the ways that local governments are interpreting data from automated decision tools in order to determine the right resources needed to combat the negative outcomes—and then make those interpretations public. Additionally, they expressed concerns about the need to broaden who is interpreting the data. Encouraging diverse perspectives that create and interpret algorithmic results is key.

CONCLUSION

The community wants to ensure that the government is addressing systemic issues; the public wants assurances that algorithms don't widen current disparities. The concern is that technology is being used as a scapegoat for erroneous decisions that people make. Algorithms run the risk of being just as biased as humans. Therefore, algorithm designers must be well-versed in racism at systematic and institutional levels. It is also essential that designers and implementers understand how algorithms have disproportionately affected people of color in the past. This will be a step in the right direction in terms of ensuring equity.
At the core of these conversations, the community wants to be assured that state and local governments truly want to strengthen and rehabilitate the community as opposed to criminalize community members. The government, in all of its dealings, needs to have the community's best interests at heart.

There is still additional work to be done to engage the community in this process. In the meantime, the task force hopes to improve general awareness of municipal algorithms and cement the public's role in the development and implementation of algorithmic systems.

> "An algorithm can be used to determine what jail you go to, what your sentence would be, determine probation and parole. This could be a chain around a Black man's neck that he can never get off because of the structure."
>
> — Community member attending public task force meeting

Endnotes

1 An exception exists in part here for facial recognition technology, which the City of Pittsburgh began to regulate through a 2020 ordinance.

2 Section III *infra* surveys some of this literature.

3 The task force's definition borrows substantially from Professor Rashida Richardson's work. Richardson offers the following "narrow" definition of automated decision systems: "any systems, software, or process that use computation to aid or replace government decisions, judgments, and/or policy implementation that impact opportunities, access, liberties, rights, and/or safety." Rashida Richardson, Defining and Demystifying Automated Decision Systems. *Maryland Law Review*. 2021:13, 14, https://ssrn.com/abstract=3811708.

4 Algorithmic Accountability Policy Toolkit (AINow), 2, https://ainowinstitute.org/aap-toolkit.pdf.

5 Richardson, Defining and Demystifying Automated Decision Systems

6 Algorithmic Accountability Policy Toolkit, 2.

7 Upturn and Omidyar Network Report, 9, https://omidyar.com/wp-content/uploads/2020/09/Public-Scrutiny-of-Automated-Decisions.pdf.

8 Upturn and Omidyar Network Report, 9.

9 For a discussion about the pitfalls of an overly broad definition (e.g., computation that assists or supplants human decision-making), see https://fpf.org/blog/automated-decision-making-systems-considerations-for-state-policymakers/.

10 DHS refers to the system as a "predictive risk model."

11 Allegheny Family Screening Tool (Allegheny County), https://www.alleghenycounty.us/Human-Services/News-Events/Accomplishments/Allegheny-Family-Screening-Tool.aspx.

12 Ibid.

13 Ibid.

14 Frequently Asked Questions, Allegheny Department of Human Services, April 2019, 7, https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/FAQs-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-8.pdf. The AFST was first tasked with making two adverse outcome predictions, although DHS eventually eliminated the second of those predictions: "The first adverse outcome predicted by the AFST is placement within two years of screen-in. Because placements are determined by a judge, and all parties (parents, children, and county) are represented by attorneys, a placement outcome is reasonably independent of the county child-welfare system." Ibid.

15 Frequently Asked Questions, Allegheny County Department of Human Services, 10.

16 Allegheny County Data Warehouse. https://www.alleghenycountyanalytics.us/index.php/dhs-data-warehouse/

17 Rhema Vaithianathan et al., Allegheny Family Screening Tool: Methodology, Version 2, Allegheny County Analytics, April 2019, 3-4, https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Methodology-V2-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-7.pdf. Vaithianathan is a professor of health economics and director of the Centre for Social Data Analytics at the University of Auckland and professor of social data analytics at the Institute for Social Science Research at the University of Queensland.

18 As part of an effort to include data from commercial insurers, the county has reportedly reached agreement with UPMC to provide data and is seeking additional agreements.

19 Allegheny County Department of Human Services Request for Proposal, https://www.alleghenycounty.us/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=2147486301, p. 11.

20 Allegheny County Family Screening Tool. Task force member Alexandra Chouldechova was involved in work to study the AFST (including its deployment, validation, and elements of its post-deployment validation) with this team. See, e.g., http://proceedings.mlr.press/v81/chouldechova18a/chouldechova18a.pdf.

21 Time Dare and Eileen Gambrill, Ethical Analysis: Predictive Risk Models at Call Screening for Allegheny County, Allegheny County Analytics, April 2017, https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Ethical-Analysis-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-2.pdf.

22 Allegheny County Department of Human Services, Ethical Analysis: Predictive Risk Models at Call Screening for Allegheny County, Allegheny County Analytics, April 2017, https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Response-to-Ethical-Analysis-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-3.pdf.

23 Frequently Asked Questions, Allegheny Department of Human Services, April 2019, 5, https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/FAQs-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-8.pdf.

24 Developing Predictive Risk Models to Support Child Maltreatment Hotline Screening Decisions, Allegheny County Analytics, April 2019, https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-2.pdf#16-ACDHS-26_PredictiveRisk_Package_041819.indd%3A.95389%3A262.

25 Ibid., Section 5

26 Ibid., Section 5.

27 Ibid.

28 Allegheny County Department of Human Services, Frequently Asked Questions, 20. "An external validation of the model was conducted using UPMC Children's Hospital of Pittsburgh data. Encounters were examined (by cause) using four approaches (highest risk score and an injury encounter, randomly selected risk score and an injury encounter, highest risk score before an injury encounter, and randomly selected risk score before an injury encounter). We found a positive correlation between the risk scores and medical encounters for injury, abusive injuries, and suicide, showing that the model accurately identifies the children most at risk for relevant hospital events." Ibid. See also Rhema Vaithianathan et al., Hospital Injury Encounters of Children Identified by a Predictive Risk Model for Screening Child Maltreatment Referrals Evidence from the Allegheny Family Screening Tool, JAMA Pediatrics. 2020, doi:10.1001/jamapediatrics.2020.2770.

29 For a critical perspective on predictive policing systems, see D. Robinson and L. Koepke, Stuck in a pattern: Early evidence on 'predictive policing' and civil rights, Upturn, August 2016, https://www.upturn.org/reports/2016/stuck-in-a-pattern/.

[30] In June 2020, the task force sent a letter to Mayor Peduto expressing concern that the program lacked transparency—for example, there was no information about the program on any city websites, and the program's development and rollout lacked consultation and engagement with the public—and that the program ought to remain suspended. Mayor Peduto responded, confirming the ongoing suspension of the program. See Letter from Pittsburgh Task Force on Public Algorithms to Mayor William Peduto, City of Pittsburgh (June 16, 2020), https://www.documentcloud.org/documents/6955009-Pittsburgh-Task-Force-on-Public-Algorithms.html#document/p1; see Letter from Mayor William Peduto, City of Pittsburgh, to Pittsburgh Task Force on Public Algorithms (June 16, 2020), https://www.documentcloud.org/documents/6955007-Pittsburgh-Task-Force-on-Public-Algorithms.html#document/p1. Mayor Peduto's letter was misdated to June 16, rather than June 20—the date that it was sent to the task force.

[31] Letter from Mayor Bill Peduto, City of Pittsburgh, to Pittsburgh Task Force on Public Algorithms (June 16, 2020), https://www.documentcloud.org/documents/6955007-Pittsburgh-Task-Force-on-Public-Algorithms.html#document/p1.

[32] Zach Goldstein, Here's how Pittsburgh cops are working to predict the future. PostIndustrial, June 30, 2019, https://postindustrial.com/featuredstories/heres-how-pittsburgh-cops-are-working-to-predict-the-future/.

[33] Dylan J. Fitzpatrick, Wilpen Gorr, and Daniel B. Neill, Hot-Spot-Based Predictive Policing in Pittsburgh: A Controlled Field Experiment, November 11, 2020, https://arxiv.org/pdf/2011.06019.pdf.

[34] Fitzpatrick, Gorr, and Neill, Hot-Spot-Based Predictive Policing in Pittsburgh, 4.

[35] Fitzpatrick, Gorr, and Neill, Hot-Spot-Based Predictive Policing in Pittsburgh, 5, 6.

[36] Goldstein, Here's how Pittsburgh cops are working to predict the future.

[37] Professor Gore's March 2021 submission to the task force included access to "experiment data" but also advised, "We do not yet have permission from the Pittsburgh Bureau of Police to share the historical data we used to develop our prediction models, but we hope to get permission in the near future." By contrast, the AFST, for example, underwent an ethical review, an impact analysis, and a process evaluation—all of which have been made available to the public—and included concerns of racial bias in the system's development.

[38] Those documents are: Dylan Fitzpatrick, Wilpen Gorr, and Daniel B. Neill, Experimental Pittsburgh Crime Hot-Spot Program: Description and Results, January 29, 2021, https://drive.google.com/file/d/18Z_2nlmpZE5RNiEwKYt8goTdFN2I-jTEq/view; Dylan J. Fitzpatrick, Wilpen Gorr, and Daniel B. Neill, Hot-Spot-Based Predictive Policing in Pittsburgh: A Controlled Field Experiment, November 11, 2020, https://arxiv.org/pdf/2011.06019.pdf; and Dylan Fitzpatrick, 2021, Replication Data for: Policing Chronic and Temporary Hot Spots of Violent Crime, https://doi.org/10.7910/DVN/HUFZ6B, Harvard Dataverse, V1, UNF:6:lUY3vo5JDG53I1Y2aQpBEQ== [fileUNF].

[39] Chamari I. Kithulgoda et al., Implementing a Predictive Risk Model to Prioritize Homeless Services: The Allegheny Homelessness Tool, Centre for Social Data Analytics, https://csda.aut.ac.nz/__data/assets/pdf_file/0005/309146/The-Allegheny-Homelessness-Tool-by-CSDA-and-ACDHSl-BloombergPoster-ForPrint.pdf.

[40] Allegheny County, Department of Human Services, Hello Baby, https://www.alleghenycounty.us/Human-Services/News-Events/Accomplishments/Hello-Baby.aspx; Hello Baby, https://hellobabypgh.org/about/.

[41] Centre for Social Data Analytics, Implementing the Hello Baby Prevention Program in Allegheny County: Methodology Report, September 2020, https://www.alleghenycountyanalytics.us/wp-content/uploads/2020/09/hello-baby-Methodology-Report_9-9-20-v2.pdf.

[42] Memorandum from Debora Daro (senior research fellow, Chapin Hall, University of Chicago) to Erin Dalton, county DHS, Hello Baby Ethical Review, April 19, 2019, https://www.alleghenycountyanalytics.us/wp-content/uploads/2020/09/Ethic-Memo-Hello-Baby_Daro_81620-edits_dd-changes.pdf; Michael Veale, Draft Ethical Report Concerning Predictive Modeling in the Allegheny Babies and Families Project, April 11, 2019, https://www.alleghenycountyanalytics.us/wp-content/uploads/2020/09/DraftEthicsReport_MVeale.pdf.

[43] Allegheny County Department of Human Services Response to Independent Reviews of the Hello Baby Predictive Risk Model and Planning Intervention and Implementation, September 2020, https://www.alleghenycountyanalytics.us/wp-content/uploads/2020/09/20-ACDHS-18-HelloBaby-CYF-Response_v2.pdf.

[44] Jackie Snow, This AI traffic system in Pittsburgh has reduced travel time by 25%, Smart Cities Dive, July 20, 2017, https://www.smartcitiesdive.com/news/this-ai-traffic-system-in-pittsburgh-has-reduced-travel-time-by-25/447494/.

[45] https://pittsburghpa.gov/press-releases/press-releases/5020.

[46] Scott Barsotti, CMU Model Is Nation's Best Fire Predictor, Carnegie Mellon University, June 28, 2018, https://www.cmu.edu/news/stories/archives/2018/june/fire-predicting-model.html.

[47] See Seyedsaeed Hajiseyedjavadi, Michael Blackhurst, and Hassan A Karimi. A Machine Learning Approach to Identify Houses With High Lead Tap Water Concentrations. Proceedings of the AAAI Conference on Artificial Intelligence. 2020;34(8):13300-305, https://doi.org/10.1609/aaai.v34i08.7040.

[48] Arnold Ventures was previously known as the John and Laura Arnold Foundation, which was its name when it launched the Public Safety Assessment.

[49] Pennsylvania Detention Risk Assessment Instrument Handbook, https://www.pachiefprobationofficers.org/docs/PaDRAI_Handbook_12-2020.pdf.

[50] Amerigo Allegretto, National, County Research Partnerships Use AI to Better Predict Opioid Overdose Risk, PittWire Health, May 22, 2019, https://www.pitt.edu/pittwire/features-articles/national-county-research-partnerships-use-ai-to-better-predict-opioid-overdose-risk; see also Lo-Ciganic et al., Evaluation of Machine-Learning Algorithms for Predicting Opioid Overdose Risk Among Medicare Beneficiaries with Opioid Prescriptions, National Center for Biotechnology Information, U.S. National Library of Medicine, and National Institutes of Health, March 1, 2019, https://www.ncbi.nlm.nih.gov/pubmed/30901048.

[51] See, e.g., Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, Cass R. Sunstein, Discrimination in the Age of Algorithms, Journal of Legal Analysis. 2018;10:113-74, https://doi.org/10.1093/jla/laz001 (arguing that algorithms "can be a potential positive force for equity"); Sendhil Mullainathan, Biased Algorithms Are Easier to Fix Than Biased People, New York Times, Dec. 9, 2019, https://www.nytimes.com/2019/12/06/business/algorithm-bias-fix.html.

[52] Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, Cass R. Sunstein, Discrimination in the Age of Algorithms, Journal of Legal Analysis. 2018;10:114, https://doi.org/10.1093/jla/laz001.

[53] Anusha Dhasarathy, Sahil Jain, and Naufal Khan, When governments turn to AI: Algorithms, trade-offs, and trust, McKinsey & Company, February 2019, https://www.mckinsey.com/industries/public-sector/our-insights/when-governments-turn-to-ai-algorithms-trade-offs-and-trust.

[54] Gretchen Greene, The Ethics of AI and Emotional Intelligence, Partnership on AI, https://www.partnershiponai.org/wp-content/uploads/2020/07/PAI_The-ethics-of-AI-and-emotional-intelligence_073020.pdf, 2.

[55] Anusha Dhasarathy, Sahil Jain, and Naufal Khan, When governments turn to AI: Algorithms, trade-offs, and trust, McKinsey & Company, February 2019, https://www.mckinsey.com/industries/public-sector/our-insights/when-governments-turn-to-ai-algorithms-trade-offs-and-trust.

[56] For an overview and general discussion of the types of problems that can plague algorithmic systems, including examples, see https://parametric.press/issue-01/the-myth-of-the-impartial-machine/; Cathy O'Neill, Weapons of Math Destruction, New York: Crown, 2016; Virginia Eubanks, Automating Inequality, New York: St. Martin's Press, 2018.

[57] See, e.g., Rashida Richardson, Jason M. Schultz, & Vincent M. Southerland, Litigating Algorithms 2019 U.S. Report: New Challenges to Government Use of Algorithmic Decision Systems, AI Now Institute, September 2019, https://ainowinstitute.org/litigatingalgorithms-2019-us.html; see Stephanie Wykstra, How the Government's Use of Algorithm Serves Up False Fraud Charges, Salon, June 15, 2020, https://www.salon.com/2020/06/14/how-the-governments-use-of-algorithm-serves-up-false-fraud-charges_partner/.

[58] Rashida Richardson, Jason M. Schultz, & Vincent M. Southerland, Litigating Algorithms 2019 U.S. Report: New Challenges to Government Use of Algorithmic Decision Systems, AI Now Institute, Septemb.er 2019, https://ainowinstitute.org/litigatingalgorithms-2019-us.html.

[59] Ibid.

[60] For a discussion of a variety of problems faced by algorithmic systems, see MITRE Partnership Network, AI Fails and How We Can Learn From Them, https://sites.mitre.org/aifails/.

[61] See, e.g., Pam Greenberg, Spotlight: Facial Recognition Gaining Measured Acceptance, National Conference of State Legislatures, September 18, 2020, https://www.ncsl.org/research/telecommunications-and-information-technology/facial-recognition-gaining-measured-acceptance-magazine2020.aspx; Claire Garvey, Garbage In, Garbage Out, https://www.flawedfacedata.com/, Georgetown Law Center on Privacy & Technology.

[62] Joy Boulamwini & Timnit Gebru, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classifications, 2018, http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf, 1, 8.

[63] Ibid., 8.

[64] Ibid., 1.

[65] https://docs.house.gov/meetings/GO/GO00/20190522/109521/HHRG-116-GO00-Wstate-BuolamwiniJ-20190522.pdf. See also Jeremy C. Fox, Brown University student mistakenly identified as Sri Lanka bombing suspect, Boston Globe, April 2019, https://www.bostonglobe.com/metro/2019/04/28/brown-student-mistaken-identified-sri-lanka-bombings-suspect/0hP2YwyYi4qrCEdxKZCpZM/story.html.

[66] ACLU Pennsylvania, Advocates Call on DA Zappala to End Use of Facial Recognition Technology. https://aclupa.org/en/news/advocates-call-da-zappala-end-use-facial-recognition-technology

[67] Ibid.

[68] Amazon, We are implementing a one-year moratorium on police use of Rekognition, June 10, 2020, https://www.aboutamazon.com/news/policy-news-views/we-are-implementing-a-one-year-moratorium-on-police-use-of-rekognition.

[69] Jeffrey Destin, Amazon Extends Moratorium on Police Use of Facial Recognition Software, Reuters, May 18, 2021, https://www.reuters.com/technology/exclusive-amazon-extends-moratorium-police-use-facial-recognition-software-2021-05-18/; Lauren Hirsch, IBM Gets Out of Facial Recognition Business, Calls on Congress to Advance Policies Tackling Racial Injustice, CNBC, June 8, 2020, https://www.cnbc.com/2020/06/08/ibm-gets-out-of-facial-recognition-business-calls-on-congress-to-advance-policies-tackling-racial-injustice.html; Jay Greene, "Microsoft Won't Sell Police Its Facial-Recognition Technology, Following Similar Moves by Amazon and IBM, Washington Post, June 11, 2020, https://www.washingtonpost.com/technology/2020/06/11/microsoft-facial-recognition/.

[70] Jason Koebler, Detroit Police Chief: Facial Recognition Software Misidentifies 96% of the Time, Vice, June 29, 2020, https://www.vice.com/en/article/dyzykz/detroit-police-chief-facial-recognition-software-misidentifies-96-of-the-time.

[71] U.S. House of Representatives, Committee on Oversight and Reform, Facial Recognition Technology (Part III): Ensuring Commercial Transparency & Accuracy, Jan. 15, 2020, https://ainowinstitute.org/oversight-committee-testimony-whittaker.pdf, 2.

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

[73] Angwin et al., Machine Bias.

[74] William Dieterich, Christina Mendoza, & Tim Brennan, COMPAS Risk Scales: Demonstrating Accuracy, Equity, and Predictive Policing, Northpointe Inc., July 8, 2016, http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf, 1.

[75] For an interesting and brief discussion of algorithmic bias (including historical underpinnings), see Kristine Gloria, Aspen Digital (Aspen Institute), Power and Progress in Algorithmic Bias, July 15, 2021, https://www.aspeninstitute.org/publications/power-progress-in-algorithmic-bias/, 2-4.

[76] Alice Feng & Shuyan Wu, The Myth of the Impartial Machine, Parametric Press, May 1, 2019, https://parametric.press/issue-01/the-myth-of-the-impartial-machine/.

[77] Ibid.

[78] Ziad Obermeyer et al., Dissecting racial bias in an algorithm used to manage the health of populations, Science, Oct. 25, 2019, 1.

[79] Ibid.

[80] Ibid.

[81] Chelsea Barabas, Ruha Benjamin, John Bowers, Meredith Broussard, et al. Technical Flaws of Pretrial Risk Assessments Raise Grave Concerns, 2019, https://cyber.harvard.edu/story/2019-07/technical-flaws-pretrial-risk-assessments-raise-grave-concerns, 1.

[82] Pittsburgh's Racial Demographics 2015: Differences and Disparities, Center on Race & Social Problems, 2015, https://www.crsp.pitt.edu/sites/default/files/2015demographics.pdf, 6.

[83] Ibid., 71.

[84] Ibid., 76.

[85] The Fourth Amendment provides that "the right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized," U.S. Constitution, Amendment IV.

[86] See generally David Casacuberta, Bias in a Feedback Loop: Fueling Algorithmic Injustice," CCAB, May 9, 2018, http://lab.cccb.org/en/bias-in-a-feedback-loop-fuelling-algorithmic-injustice/.

[87] Feng & Wu, The Myth of the Impartial Machine.

[88] "In machine learning models where the prediction is fed back into the model as inputs for the next round of predictions, bias can be amplified further in the form of a feedback loop." Feng and Wu, The Myth of the Impartial Machine.

[89] https://www.dictionary.com/browse/feedback-loop. See also Nikhil Sonnad, Data scientist Cathy O'Neil on the cold destructiveness of big data, Quartz, Dec. 7, 2016, https://qz.com/819245/data-scientist-cathy-oneil-on-the-cold-destructiveness-of-big-data/.

[90] Rashida Richardson, Jason Schultz, and Kate Crawford, Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice, February 13, 2019. 94 N.Y.U. L. REV. ONLINE 192, 193, https://ssrn.com/abstract=3333423.

[91] Neil Bedi & Kathleen McGrory, Pasco's sheriff uses grades and abuse histories to label schoolchildren potential criminals, Tampa Bay Times, Nov. 19, 2020, https://projects.tampabay.com/projects/2020/investigations/police-pasco-sheriff-targeted/school-data.

[92] To be clear, in making data publicly available as part of any effort to scrutinize a system, sensitive personal identifying information should not be made public, and privacy-preserving mechanisms should be used.

[93] Robert Brauneis & Ellen P. Goodman, Algorithmic Transparency for the Smart City, Yale Journal of Law and Technology. 2018;103:119.

[94] Ibid., 110.

[95] Ibid., 120.

[96] Ibid., 120.

[97] See West SM, Whittaker M, and Crawford K, Discriminating Systems: Gender, Race and Power in AI. AI Now Institute, 2019: 10-12, https://ainowinstitute.org/discriminatingsystems.html.

[98] See, e.g., Emily Adlin Bosk, What counts? Quantification, worker judgment, and divergence in child welfare decision making, Human Services Organizations: Management, Leadership & Governance. 2018;42(2):205-224, doi: 10.1080/23303131.2017.1422068; Emily A. Bosk, Iron Cage or Paper Cage? The Interplay of Worker Characteristics and Organizational Policy in Shaping Unequal Responses to a Standardized Decision-Making Tool, Social Problems, 2019, doi: 10.1093/socpro/spz028.

[99] See John Howard et al., Human-algorithm teaming in face recognition: how algorithm outcomes cognitively bias human decision-making, PloS One, 2020;15(8):e0237855, doi:10.1371/journal.pone.0237855.

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

[101] Centre for Data Ethics and Innovation, Review into Bias in Algorithmic Decision-Making (November 2020).

[102] https://www.theregreview.org/2019/02/12/goodman-smart-algorithmic-change-requires-collaborative-political-process/.

[103] Ibid.

[104] In a related vein, work surrounding government's "license to operate" algorithmic systems for the public reinforces the pitfalls where such systems are not transparent or accountable to the public. Shah H, Algorithmic accountability, Philosophical Transactions of the Royal Society A, 2018;376:20170362, http://dx.doi.org/10.1098/rsta.2017.0362.

[105] See, e.g., New York City Automated Decision Systems Task Force Report, November 2019, https://www1.nyc.gov/assets/adstaskforce/downloads/pdf/ADS-Report-11192019.pdf; Considerations for Implementing Predictive Analytics in Child Welfare, April 2018, Casey Family Programs, https://caseyfamilypro-wpengine.netdna-ssl.com/media/Considerations-for-Applying-Predictive-Analytics-in-Child-Welfare.pdf; Dillon Reisman, Jason Schultz, Kate Crawford, & Meredith Whittaker, Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability, AI Now Institute, April 2018, https://ainowinstitute.org/aiareport2018.pdf; Algorithm Charter for Aotearoa New Zealand, https://data.govt.nz/use-data/data-ethics/government-algorithm-transparency-and-accountability/algorithm-charter; CDEI Review into Bias in Algorithmic Decision-Making, November 2020, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957259/Review_into_bias_in_algorithmic_decision-making.pdf; Kristine Gloria, Aspen Digital, Aspen Institute, Power and Progress in Algorithmic Bias, July 15, 2021, https://www.aspeninstitute.org/publications/power-progress-in-algorithmic-bias/; Open Government Partnership, Algorithmic Accountability for the Public Sector: Learning From the First Wave of Policy Implementation, March 24, 2021, https://www.opengovpartnership.org/stories/algorithmic-accountability-for-the-public-sector-learning-from-the-first-wave-of-policy-implementation/.

[106] Gretchen Greene, Potholes, Rats and Criminals: A Framework for AI Ethical Risk, Data-Smart City Solutions, Harvard Kennedy School, April 19, 2018, https://datasmart.ash.harvard.edu/news/article/potholes-rats-and-criminals. For discussion of a similar potential risk-based framework in the European Union, see https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/what-the-draft-european-union-ai-regulations-mean-for-business.

[107] Gretchen Greene, Potholes, Rats and Criminals: A Framework for AI Ethical Risk, Data-Smart City Solutions, Harvard Kennedy School, April 19, 2018, https://datasmart.ash.harvard.edu/news/article/potholes-rats-and-criminals.

[108] Anna Brown, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin, and Rhema Vaithianathan, Toward Algorithmic Accountability in Public Services, Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI 19, May 2019, doi:10.1145/3290605.3300271, 2.

[109] Data Futures Partnership, 2017, A Path to Social License: Guidelines for Trusted Data Use, https://aisp.upenn.edu/wp-content/uploads/2019/08/Trusted-Data-Use_2017.pdf.

[110] Shah H. 2018 Algorithmic accountability. Phil. Trans. R. Soc. A, 376:20170362, http://dx.doi.org/10.1098/rsta.2017.0362.

[111] Ibid.

[112] See Anna Brown, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin, and Rhema Vaithianathan. Toward algorithmic accountability in public services. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI 19, May 2019, doi:10.1145/3290605.3300271, 2

[113] A significant change in technical metrics, data, or process, for example, should trigger the stakeholder to bring in the public.

[114] Dillon Reisman, Jason Schultz, Kate Crawford, & Meredith Whittaker, Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability, AI Now Institute, April 2018, https://ainowinstitute.org/aiareport2018.pdf at 5.

[115] Independence would be key (i.e., those with a vested interest in the system under consideration should be avoided).

[116] Tim Dare and Eileen Gambrill, Ethical Analysis: Predictive Risk Models at Call Screening for Allegheny County, Allegheny County Analytics, April 2017, https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Ethical-Analysis-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-2.pdf (ethical review); Developing Predictive Risk Models to Support Child Maltreatment Hotline Screening Decisions, Allegheny County Analytics, April 2019, https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-2.pdf#16-ACDHS-26_PredictiveRisk_Package_041819.indd%3A.95389%3A262 (process evaluation); Ibid., Section 5 (impact evaluation).

[117] It will be important to include in a process review an understanding of how the system was planned and later used during deployment.

[118] City of Pittsburgh Department of Innovation and Performance Data Governance Operational Charter.

[119] Pitt Cyber plans to publish a website prototype for consideration in 2022.

[120] ACLU Pennsylvania, Advocates Call on DA Zappala to End Use of Facial Recognition Technology, Oct. 21, 2019, https://www.aclupa.org/en/news/advocates-call-da-zappala-end-use-facial-recognition-technology.

[121] Juliette Rihl, 'If Your Mom Can Go in and See It, So Can the Cops': How Law Enforcement Is Using Social Media to Identify Protestors in Pittsburgh," PublicSource, Aug. 6, 2020, https://www.publicsource.org/pittsburgh-police-arrest-blm-protesters-social-media-facial-recognition/.

[122] Ibid.

[123] Ashley Murray, Pittsburgh City Council Approves Bill to Regulate Facial Recognition Technology, Pittsburgh Post-Gazette, Sept. 22, 2020, https://www.post-gazette.com/news/politics-local/2020/09/22/Pittsburgh-City-Council-approves-bill-to-regulate-facial-recognition-technology/stories/202009220113.

[124] https://apps.pittsburghpa.gov/redtail/images/14012_FINAL_DRAFT_7_Annual_Report_2020.pdf (four instances in 2020); https://apps.pittsburghpa.gov/redtail/images/9640_2019_Annual_Report_Final.pdf (nine instances in 2019).

[125] Robert Brauneis & Ellen P. Goodman, Algorithmic Transparency for the Smart City, Yale Journal of Law and Technology. 2018; 103:116.

University of Pittsburgh

Institute for Cyber Law, Policy, and Security

4200 Fifth Avenue
2516 Cathedral of Learning
Pittsburgh, PA 15260

cyber.pitt.edu