ALGORITHMIC TRANSPARENCY FOR THE SMART CITY

by Robert Brauneis and Ellen P. Goodman*

"As a society, we are now at a crucial juncture in determining how to deploy AI-based technologies in ways that promote, not hinder, democratic values such as freedom, equality, and transparency."[1]

ABSTRACT

Emerging across many disciplines are questions about algorithmic ethics – about the values embedded in artificial intelligence and big data analytics that increasingly replace human decisionmaking. Many are concerned that an algorithmic society is too opaque to be accountable for its behavior. An individual can be denied parole or denied credit, fired or not hired for reasons she will never know and cannot be articulated. In the public sector, the opacity of algorithmic decisionmaking is particularly problematic both because governmental decisions may be especially weighty, and because democratically-elected governments bear special duties of accountability. Investigative journalists have recently exposed the dangerous impenetrability of algorithmic processes used in the criminal justice field – dangerous because the predictions they make can be both erroneous and unfair, with none the wiser.

We set out to test the limits of transparency around governmental deployment of big data analytics, focusing our investigation on local and state government use of predictive algorithms. It is here, in local government, that algorithmically-determined decisions can be most directly impactful. And it is here that stretched agencies are most likely to hand over the analytics to private vendors, which may make design and policy choices out of the sight of the client agencies, the public, or both. To see just how impenetrable the resulting "black box" algorithms are, we filed 42 open records requests in 23 states seeking essential information about six predictive algorithm programs. We selected the most widely-used and well-reviewed programs, including those developed by for-profit companies, nonprofits, and academic/private sector

* Robert Brauneis is Professor of Law at the George Washington University Law School; Ellen P. Goodman is Professor of Law at Rutgers Law School. We would like to thank Erin Dalton, Jeremy Heffner, Andrew Nicklin, and the participants in the University of Cambridge conference on The Power Switch: How Power is Changing in a Networked World, MetroLab Network's workshop on Ethical Guidelines for Applying Predictive Tools within Child Welfare Services, the Bloomberg Philanthropies What Works Cities Summit, the Wharton School's Law and Ethics of Big Data Colloquium, and the 18th Annual Congress of the European Intellectual Property Institutes Network.

[1] Stanford University, One Hundred Year Study on Artificial Intelligence (AI100), August 1, 2016, https://ai100.stanford.edu.

partnerships. The goal was to see if, using the open records process, we could discover what policy judgments these algorithms embody, and could evaluate their utility and fairness.

To do this work, we identified what meaningful "algorithmic transparency" entails. We found that in almost every case, it wasn't provided. Over-broad assertions of trade secrecy were a problem. But contrary to conventional wisdom, they were not the biggest obstacle. It will not usually be necessary to release the code used to execute predictive models to dramatically increase transparency. We conclude that publicly-deployed algorithms will be sufficiently transparent only if (1) governments generate appropriate records about their objectives for algorithmic processes and subsequent implementation and validation; (2) government contractors reveal to the public agency sufficient information about how they developed the algorithm; and (3) public agencies and courts treat trade secrecy claims as the limited exception to public disclosure that the law requires. We present what we believe are eight principal types of information that records concerning publicly implemented algorithms should contain.

INTRODUCTION

With ever greater frequency, governments are using computer algorithms to conduct public affairs. This is especially true in cities, counties and states, whose governments are tasked with providing basic services and deploying coercive police power. The "smart city" movement worldwide impresses on local governments the importance of gathering and deploying data more effectively.[2] One of the goals is to find patterns in big data sets – for example, the places and times crime is most likely to occur - and to generate predictive models to guide the allocation of public services – for example, how and where to police.[3] Most local governments lack the expertise and wherewithal to deploy data analytics on their own. If they want to be "smart," they need to contract with companies, universities, and nonprofits to implement privately-developed algorithmic processes. The result is that privately developed predictive algorithms are shaping local government actions in such areas as criminal justice, food safety, social services, and transportation.[4]

Because the designing entities typically do not disclose their predictive models or algorithms, there is a growing literature criticizing the "black box" opacity of these processes.[5] These black boxes are impervious to question, and many worry that they

---

[2] *See, e.g.*, Rob Kitchin, The Real-Time City? Big Data and Smart Urbanism, 79 GEOJOURNAL 1 (2014).

[3] *See* Elizabeth E. Joh, The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing, 10 Harv. L. & Pol'y Rev. 15, 38 (2016); Andrew Guthrie Ferguson, *Policing Predictive Policing* . 94 WASH. U. L. REV. ___ (forthcoming 2017), https://ssrn.com/abstract=2765525; ANDREW GUTHRIE FERGUSON, THE RISE OF BIG DATA POLICING: SURVEILLANCE, RACE, AND THE FUTURE OF LAW ENFORCEMENT (2017).

[4] On smart-city algorithms generally, *see infra* notes 31-33; on the algorithms about which we filed open records requests, *see infra* pp. 26-38.

[5] *See, e.g.*, CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION (2016); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015); Jenna Burrell, How the Machine 'Thinks:' Understanding Opacity in Machine Learning Algorithms, 3 *Big Data &Society* 1 (2016), https://ssrn.com/abstract=2660674; Kate Crawford, Can an Algorithm be Agonistic? Ten Scenes from Life in Calculated Publics, *Science, Technology & Human Values* 41(1): 77–92 (2016); Tarleton Gillespie, The Relevance of Algorithms, in Media Technologies: Essays on Communication, Materiality, and Society 167-193 (Tarleton Gillespie et al. eds., 2014); Rob Kitchin, Thinking Critically about and Researching Algorithms, *Information, Communication & Society* 20(1): 14–29 (2016); Rob Kitchin, The Data Revolution: Big Data, Open Data, Data Infrastructures and their Consequences (2014); Taina Bucher, 'Want to be on the top?' Algorithmic Power and the Threat of Invisibility on Facebook, *New Media and Society*, 14(7), 1164–1180 (2014); David Beer, The Social Power of Algorithms, 20 *J. of Info., Comm., & Soc.* 1(2016); Michael Ananny, Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness, *Science, Technology & Human Value*s, 41(1), 93–117 (2015); Danielle Citron & Frank Pasquale, The Scored Society: Due Process for Automated Predictions, 89 WASH. LAW REV. 1 (2014); Nicholas Diakopoulos, Algorithmic Accountability: Journalistic Investigation of Computational Power Structures, *Digital Journalism*, 3(3), 398–415 (2015) (hereinafter "Algorithmic accountability"); Christian Sandvig, Seeing the Sort: The Aesthetic

may be discriminatory,[6] erroneous, or otherwise problematic.[7] Journalists and scholars who have begun to seek details from public entities about these algorithms generally come up short as their freedom of information requests are denied or go unanswered.[8]

Commentators have called for more transparency across all implementations of artificial intelligence.[9] There are special concerns when municipal and other governments use predictive algorithms whose development and implementation neither the public nor the government itself really understands. By developing and selling these systems to government – or even giving them away – private entities assume a significant role in public administration. What is smart in the smart city

---

and Industrial Defense of "The Algorithm," *Journal of the New Media Caucus*, 1–21 (2015); Christian Sandvig, et al., Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms (2014), available at http://www-personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf; Malte Ziewitz, Governing Algorithms: Myth, Mess, and Methods, 41 *Sci., Tech. & Human Values* 3–16 (2015); Zynep Tufekci, *Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency*, 13 J. OF TELECOM. AND HIGH TECH. LAW 203 (2015); *see generally* Tarleton Gillespie, & Nick Seaver, Critical Algorithm Studies: A Reading List, http://socialmediacollective.org/reading-lists/critical-algorithm-studies/.

[6] *See, e.g.*, Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671 (2016); Federal Trade Commission, Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues, Jan. 2016, https://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report; Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, Machine Bias, ProPublica, May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (evaluating algorithmic risk assessments used by judges to set bail amounts in Ft. Lauderdale, FL); Zeynep Tufecki, *Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency*, 13 COLO. TECH. L.J. 203 (2015).

[7] *See generally*, Lee Rainie & Janna Anderson, Code-Dependent: Pros and Cons of the Algorithm Age, Pew Reports (Feb. 9, 2017), http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/.

[8] *See, e.g.*, Nicholas Diakopoulos, We Need to Know the Algorithms the Government Uses to Make Important Decisions About Us, *The Conversation* (May 23, 2016), https://theconversation.com/we-need-to-know-the-algorithms-the-government-uses-to-make-important-decisions-about-us-57869 (reporting on open records requests to fifty states on their use of algorithms in criminal justice, of which nine "based their refusal to disclose details about their criminal justice algorithms on the claim that the information was really owned by a company."); Tonia Hill, Jamie Kalven Joins Other Chicago Journalists in Lawsuit Against CPD, *Hyde Park Herald* (June 7, 2017), http://hpherald.com/2017/06/07/jamie-kalven-joins-chicago-journalists-lawsuit-cpd/ (journalists suing Chicago Police Department for withholding information about an algorithm that produces a Strategic Subject List, known as a "heat list," predicting people allegedly likely to be involved in gun violence).

[9] *See, e.g.*, Future of Life Institute, 23 Principles for Beneficial Artificial Intelligence, https://futureoflife.org/ai-principles/ (Jan. 17, 2017) ((more than 1,600 signatories, including Steven Hawking, Elon Musk, and AI researchers called for "Failure Transparency" showing why an AI system might have caused harm and "Judicial Transparency" providing a satisfactory explanation auditable by a competent human authority of any judicial decision).

comes to reside in the impenetrable brains of private vendors while the government, which alone is accountable to the public, is hollowed out, dumb and dark. The risk is that the opacity of the algorithm enables corporate capture of public power. When a government agent implements an algorithmic recommendation that she does not understand and cannot explain, the government has lost democratic accountability, the public cannot assess the efficacy and fairness of the governmental process, and the government agent has lost competence to do the public's work in any kind of critical fashion.

We set out to test just how opaque local government predictive algorithms are. We identified the most common local government uses of big data prediction. We then assembled a "portfolio" of open records requests targeting a variety of uses and jurisdictions. We identified algorithms developed by foundations, private corporations, and government entities and those used in criminal justice and in civil applications. Using Muckrock, the nonprofit collaborative platform for filing open records requests,[10] we filed 42 requests in 23 states for records relating to six predictive algorithms.[11] The federal government and all fifty states (and Washington D.C.) have open records laws that require varying amounts of disclosure concerning the public use of algorithms. Given how broadly most open records acts are written, contracts and related correspondence with vendors will almost always be "public records" that must be disclosed.[12] Software is a "record" disclosable under the federal Freedom of Information Act (FOIA),[13] as well as under many state laws, but not all.[14] We sought records including correspondence, contracts, software, training materials, existing and planned validation studies and other documentation.

---

[10] https://www.muckrock.com/. In one case (Allegheny County Child and Family Services), we filed the requests separately, not using the platform.

[11] Our project page on Muckrock can be found at https://www.muckrock.com/project/uncovering-algorithms-84/. That page links our requests, and most of the documents provided in response to our requests. (Some governments provided links to files on their servers, rather than uploading the documents to Muckrock.) Some of our requests were initially routed to the wrong agencies; we are not counting those in the numbers we provide in the text, but they are included on the Muckrock project page.

[12] *See, e.g.*, Fla. Stat. § 119.011(12) (A "public record" open for inspection "means all documents, papers, letters, maps, books, tapes, photographs, films, sound recordings, data processing software, or other material, regardless of the physical form, characteristics, or means of transmission, made or received pursuant to law or ordinance or in connection with the transaction of official business by any agency.").

[13] *See generally*, Katherine Fink, *Opening the Government's Black Boxes: Freedom of Information and Algorithmic Accountability*, Information, Communication & Society, DOI: 10.1080/1369118X.2017.1330418 (2017) (research on federal agency responses to FOIA requests for source code).

[14] *Compare* Ark. Code Ann. § 25-19-103(5)(B) (the definition of a "public record" does not include "software acquired by purchase, lease, or license.") with Fla. Stat. § 119.011(1) ("Data processing software" is included in the definition of a "public record."). *See generally infra* notes 106-110.

What we learned is that there 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. In this article, we investigate each of these impediments, and suggest policies and practices to lower them. If these problems were addressed, we suspect that in some cases, there would be yet another impediment to real transparency: the use of algorithms that are highly dynamic or that use modeling that makes them difficult to interpret even when records are revealed. We save this issue for another day.

In Part I of this Article, we review local government use of predictive algorithms and associated big data analytics. Such use raises questions about the politics embedded in these programs, and about their utility, fairness, impact on governmental capacity, and relationship to government transparency values. Part II describes the open records requests we submitted to various jurisdictions about their deployment of predictive algorithms, and the responses we received. Part III identifies obstacles to greater transparency with respect to algorithmic processes and Part IV suggests mitigation techniques to maximize algorithmic transparency. Part V concludes.

I. Algorithmic Governance at the Local Level

Use of big data and predictive algorithms is a form of governance – that is, a way for authorities to manage individual behavior and allocate resources.[15] It consists of capturing data from multiple sources, applying data analytics to find correlations between characteristics and outcomes, and using this analysis to generate predictions that may not be easily explained or understood.[16] When algorithms are deployed in the public sphere, public authority typically yields to the private control of technology companies and other developers.[17] We start with a description of algorithmic

[15] Marijn Janssen & George Kuk, The Challenges and Limits of Big Data Algorithms in Technocratic Governance, *Government Information Quarterly* 33 (2016) 371–377 (discussing how algorithms and big data become a form of governance, often impervious to interrogation or explanation).

[16] *See* Antoinette Rouvroy & Thomas Berns (trans. Elizabeth Libbrecht), Algorithmic Governmentality and Prospects of Emancipation, Disparateness as a precondition for individuation through relationships? Réseaux 1(177): 163–96 (2013) (drawing on Foucoult's conceptions of governance to break down algorithmic governance). *See also* Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo,Sandra Wachter & Luciano Floridi, The Ethics of Algorithms: Mapping the Debate, *Big Data & Society*, July–December 2016: 1–21 at 4-5 (identifying epistemic concerns with the opacity of data inputs and processing).

[17] Rouvroy & Berns, *supra* note 16, at 5 (there is a "colonization of public space by a hypertrophied private sphere"). Aneesh Aneesh has posited an era of "algocratic governance" that supplants "bureaucratic hierarchies." Aneesh Aneesh, Technologically Coded Authority:

governance and proceed to the role of private vendors in rolling out these programs in "smart cities." We then proceed to identify the kinds of questions the public needs answered about algorithmic governance and the existing government transparency tools that might serve in this pursuit.

A. The Promise of Algorithmic Governance

An algorithm is a set of "encoded procedures for transforming input data into a desired output, based on specified calculations."[18] Like a recipe, it provides instructions for transforming ingredients into a simple or complex product.[19] All algorithms share the trait of formalization. Unlike human decisionmaking, which cannot be fully articulated or discovered, the calculations embedded in algorithms can in theory always be fully described, and unless they intentionally incorporate randomness, should yield reproducible results.[20]

Predictive algorithms, which are increasingly used in smart-city applications,[21] are created through analysis of large datasets, typically with the aid of machine-learning

---

The Post-Industrial Decline in Bureaucratic Hierarchies, http://web.stanford.edu/class/sts175/NewFiles/Algocratic%20Governance.pdf (theorizing a transition from bureaucratic governance to distributed architecture of information systems where control is exercised through code).

[18] Tarleton Gillespie, The Relevance of Algorithms, in *Media Technologies: Essays on Communication, Materiality and Society* 167, 167 (Tarleton Gillespie, Pablo J. Boczkowski, Kirsten A. Foot, eds. 2014).

[19] Algorithmic processes may be simple and the data inputs limited. For example, an algorithm could aggregate complaints about potholes and rank the potholes for repair priority based on the number of complaints. Another algorithm may encode simple procedures, but operate on a broader set of data. For example, it could allocate points for diet, exercise, and sleep, and then produce a fitness score by combining those data pointswith attributes like age, ethnicity, and location. Other algorithmic processes may be more complex, for example, looking for correlations between fitness attributes and longevity or health data among relevant populations.

[20] In spite of their formality, algorithms can of course embody complex human judgments and contested categories. For example, an algorithm can assign a precise weight to the race of an individual to predict some behavior, but the decisions about how to create racial categories, and how to assign each individual to one of those categories, still depends on human judgment that is not itself formalized, and may be highly controversial. For discussion of the HunchLab algorithm's incorporation of randomness, *see infra* p. 47.

[21] Predictive algorithms can be distinguished from other kinds of algorithms governments use to encode publicly revealed and specified policy rules – rules derived from laws and regulations that do not require further specification or development. For example, state governments use algorithms to determine eligibility for benefit programs such as Food Stamps. Applicants submit facts about themselves – the input – and a computer applies a set of rules – the specified calculation – to produce an output: a determination of "eligible" or "not eligible." In that case, the algorithm consists of a series of conditional statements, and the output is an answer to a yes/no question. The policy rules might also accept quantitative inputs, such as the age and number of people in a household, and produce quantitative outputs, such as the level of government benefits that will be paid to that household.

processes,[22] to reveal correlations between various features (of a person, circumstance, or activity) and desired or objectionable outcomes.[23] Those patterns can then be used to create a model that will estimate the likelihood of future behavior or events (the output) when given relevant facts (the input).[24] An algorithmic process will therefore typically involve (i) the construction of a model to achieve some goal, based on analysis of collected historical data, (ii) the coding of an algorithm that implements this model, (iii) collection of data about subjects to provide inputs for the algorithm, (iv) application of the prescribed algorithmic operations on the input data, and (v) outputs in the form of predictions or recommendations based on the chain of data analysis.[25]

For example, a government may want to know how likely a prisoner is to commit a crime if paroled, or how likely an admitted student is to enroll in a state university if offered a scholarship of a certain amount. By correlating a set of characteristics of past parolees with their subsequent criminal histories, or of past admitted students with their enrollment decisions, data scientists can build a predictive model. The government can then apply that model to current parolees or admitted students, and predict their behavior.

Governmental deployment of algorithmic processes promises increased efficacy and fairness in the delivery of government services. Data analysis can surface patterns not previously noticed or not precisely quantified. For example, systematic tracking of Yelp restaurant reviews can inform city health inspectors about food-borne illnesses emerging from the restaurants in their jurisdictions.[26] Integrating data from across siloed administrative domains, such as education and human services, and then using

---

[22] For a definition, see Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C L. REV. 93, 96 (2014) ("the term encompasses three aspects of data magnification and manipulation. First, it refers to technology that maximizes computational power and algorithmic accuracy. Second, it describes types of analyses that draw on a range of tools to clean and compare data. Third, it promotes the belief that large data sets generate results with greater truth, objectivity, and accuracy.") (omitting citations).

[23] *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights, Executive Office of the President* (May 2016) at p. 8 (machine learning is the "'science of getting computers to act without being explicitly programmed.'")(quoting Ng, A. Coursera Machine Learning course. Stanford University 2016).

[24] Robin K. Hill, What an Algorithm Is, 29(1) *Philosophy &Technology* 35–59 (2015).

[25] See Tal Zarsky, Transparent Predictions, 2013 U. Ill. L. Rev. 1503, 1517-1520; Gillespie, *supra* note 5, at 167; Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson & Harlan Yu, Accountable Algorithms, 165 U. Penn. L. Rev. 633, 640 n. 14 (2017)

[26] *See* Edward L. Glaeser, et al., Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life, *Harvard Business School NOM Unit Working Paper 16-065*, 2015, https://dash.harvard.edu/bitstream/handle/1/24009688/16-065.pdf?sequence=1.

that data to prioritize families in need of government help, can improve social service delivery.[27]

Algorithmically-informed decisionmaking can also help government officials avoid the biases, explicit or implicit, that may creep into less formal, "hunch"-based decisionmaking. [28] For example, members of a parole board who simply interview a prisoner to make parole decisions may be overly focused on the severity of the crime, or on whether the prisoner displays remorse, or on cultural or ethnic generalizations. By contrast, the systematic use of data analytics can identify characteristics that have a significant correlation with recidivism and evaluate the strength of those correlations, either separately or in combination. Those correlations can then be encoded into an algorithm that estimates of the risk of recidivism when fed input information about the prisoner.[29]

B. SMART CITIES EN MARCHE

Implementation of algorithms at the local level is part of a broader move towards data-driven decisionmaking, and must be understood in the context of the smart city agenda. In the 21st century, cities and counties have increasingly turned to "digital hardware and software, producing massive amounts of data about urban processes."[30] At first, the integration of digital technologies into governance involved rudimentary e-government initiatives and digitizing governmental resources.[31] In the past half-decade, local governments have turned towards more extensive analytics and the exploitation of sensor networks, ubiquitous communications, and computing.[32]

Smart city initiatives seek to harness data to rationalize and automate the operation of public services and infrastructure, such as transportation, energy, and health services.[33]

---

[27] *See, e.g.*, Erika M. Kitzmiller, IDS Case Study: Allegheny County's Data Warehouse: Leveraging Data to Enhance Human Service Programs and Policies. Actionable Intelligence for Social Policy (AISP), University of Pennsylvania (2013) (analyzing how Allegheny County Pennsylvania has used data analytics to improve its human service agency's responsiveness).

[28] *See* Daniel Castro (2016) Data detractors are wrong: The rise of algorithms is a cause for hope and optimism. Center for Data Innovation, https://www.datainnovation.org/2016/10/data-detractors-are-wrong-the-rise-of-algorithms-is-a-cause-for-hope-and-optimism/.

[29] On the Arnold Foundation's PSA-Court algorithms, concerning which we filed open records requests, *see infra* pp. 26-30.

[30] Alan Wiig & Elvin Wyly, Introduction: Thinking Through the Politics of the Smart City, *Urban Geography*, 37:4, 485-493 at 488 (2016). *See also*, Rob Kitchin, The Real-Time City? Big Data and Smart Urbanism, 79 *GeoJournal* 1 (2014).

[31] *See generally* STEPHEN GOLDSMITH & SUSAN CRAWFORD, THE RESPONSIVE CITY: ENGAGING COMMUNITIES THROUGH DATA-SMART GOVERNANCE (2014).

[32] *See generally* ANTHONY M. TOWNSEND, TOWNSEND, SMART CITIES: BIG DATA, CIVIC HACKERS, AND THE QUEST FOR A NEW UTOPIA (2013).

[33] *See* Lilian Edwards, *Privacy, Security and Data Protection in Smart Cities: A Critical EU Law Perspective*, 2 EUR. DATA PROTECTION L. REV. 28-58, (2016), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2711290;

Cities are being asked to handle more with fewer resources as they transition to data-based governance. They can't do it without public-private partnerships, which develop the analytics and ensuing "smart" systems.[34] Private entities have been at the leading edge of the entire smart city movement.[35] Indeed, IBM registered the phrase "smarter cities" as a trademark as part of its campaign to market technology-driven urban management.[36] Cisco has been similarly active.[37] It is to these companies and other private vendors that local government officials, pressed by economic necessity and personnel constraints, will often leave the work of data analytics. Lillian Edwards reports that governments usually cede ownership of the underlying data as well to their private partners. "Policing, surveillance, crowd control, emergency response, are all historically state functions, and citizens might expect the very sensitive data involved to be held by the state. Yet the likelihood in a . . . city [built on public-private partnerships] is that the data finds itself . . . in private control."[38]

In the case of predictive algorithms, the vendor often presents the government with a standard contract giving the vendor control and/or ownership of the data and the analytics.[39] Some smart city commentators warn that "smart" projects are simply vehicles to sell municipalities comprehensive data management systems owned and

---

Hafedh Chourabi, *et al.*, Understanding Smart Cities: An Integrative Framework. In *Proceedings of the 2012 45th Hawaii International Conference on System Sciences* (HICSS 2012), available at http://dx.doi.org/10.1109/HICSS.2012.615 (describing and synthesizing various conceptions of the smart city). *See also*, Nils Walravens and Pieter Ballon, Platform Business Models for Smart Cities: From Control and Value to Governance and Public Value, *51 Communications Magazine, IEEE, 6* (2013)(discussing role of mobile technologies in addressing urban problems); Ellen P. Goodman, *"Smart Cities" Meet "Anchor Institutions": The Case for Broadband and the Public Library*, 41 FORDHAM URB. L.J. 1665 (2015).

[34] *See, e.g.*, Alberto Vanolo, *Smartmentality: The Smart City as Disciplinary Strategy*, Urban Studies 51(5) 883-898 (2013) (critically describing the centrality of public-private partnership to the smart city vision and implementation).

[35] *See* Janine S. Hiller & Jordan M. Blanke, *Smart Cities, Big Data, and the Resilience of Privacy* (forthcoming HASTINGS L. J.) (describing the corporate framing of smart cities)

[36] *See* U.S. Trademark Registration No. 4033245 (issued October 4, 2011); Ola Söderström, et al, Smart Cities as Corporate Storytelling, *City*, Vol. 18: 307-320 (2014); *see also* A. Wiig, IBM's Smart City as Techno-Utopian Policy Mobility, *City*, Vol. 19: (2015); Wiig, *supra* note xx at 540 [2016] (IBM's Smarter Cities Challenge offered cities partnerships with corporate "consultants and technology specialists will help municipalities analyze and prioritize their needs, review strengths and weaknesses, and learn from the successful strategies used by other cities.").

[37] Gordon Falconer & Shane Mitchell, Cisco, Smart City Framework: A Systematic Process of Enabling Smart + Connected Communities, http://www.cisco.com/c/dam/en_us/about/ac79/docs/ps/motm/Smart-City-Framework.pdf (2012).

[38] Edwards, *supra* note 33, at 33.

[39] *See, e.g., infra* note 118; *see also* Angwin et al., *supra* note 6 (describing the COMPAS contract).

managed by the vendor.[40] The fear is that smart city partnerships will ultimately lead to the surrender of public services to private interests. Service contracts can make governments dependent on the technology provider for upgrades and ongoing development, locking the government into proprietary technologies whose costs and pace of innovation they can't control. A related concern is that the private vendor comes to own critical data. According to the digital chief of Barcelona, a leader in smart city technologies, cities can "end up with a black-box operating system where the city itself loses control of critical information and data that should be used to make better decisions."[41] The risk is that the corporation controlling the data and analytics occupies the command center of urban governance while the democratically accountable officials move to the periphery.[42]

C. What the Public Needs to Know

Algorithmic governance has a politics. When private vendors control algorithmic governance, the politics of algorithms recede behind private hedges. In other areas of privatization – schools and prisons – the stakes are clear. It is less obvious what is at stake with private control of algorithmic governance in part because algorithms may seem like science without a politics. Algorithms positioned merely as the means to scientific truth can conceal the values embedded in the underlying models.[43] Yet

---

[40] *See, e.g.*, Adam Greenfield, Against the Smart City (2013); Donald McNeill, Global firms and Smart Technologies: IBM and the Reduction of Cities, *Transactions of the Institute of British Geographers*, 40(4), 562–574 (2015); Alan Wiig, The Empty Rhetoric of the Smart City: From Digital Inclusion to Economic Promotion in Philadelphia, *Urban Geography*, 37:4, 535-553 (2016) ("the smart city acts as a data-driven logic urban change where widespread benefit to a city and its residents is proposed, masking the utility of these policies to further entrepreneurial economic development strategies").

[41] David Meyer, How One European Smart City is Giving Back Power to its Citizens, http://www.alphr.com/technology/1006261/how-one-european-smart-city-is-giving-power-back-to-its-citizens (Jul. 10, 2017) (quoting Francesca Bria). Another problem is that "the business model is creating dependence on very few providers." This lock-in of city services to particular private vendors could "be extended to the entire urban infrastructure of the city. We're talking about transportation, better waste management, even water, energy, distributed green infrastructure. It's a big problem for a public administration, losing control of the management of the infrastructure." *Id.*

[42] Christine Richter & Linnet Taylor, *Geographies of Urban Governance* 175-191 in J. Gupta et al., eds, Big Data and Urban Governance (2015) at 180 ("The increasing influence of corporations over the creation of the smart city environment potentially places corporations at the centre of democratic urban processes."). *See also* Kitchin, *supra* note xx [The real-time city? Big data and smart urbanism. GeoJournal 79(1):1–14].

[43] *See* Rob Kitchen, Reframing, Reimagining and Remaking Smart Cities, *The Programmable City* at 4 (2016) (summarizing smart city critiques).

judgments are encoded in the algorithmic process at all stages.[44] These are judgments that at some level the public should know and speak to.

1. What Are the Algorithm's Politics?

As Harry Surden notes, a predictive algorithm's recommendation "actually masks an underlying series of subjective judgments on the part of the system designers about what data to use, include or exclude, how to weight the data, and what information to emphasize or deemphasize."[45] There will be tradeoffs in implementing any policy goal, even one as uncontroversial as reducing traffic wait time. What risk to pedestrian safety is permissible in the service of traffic flow? How does the reduction of tailpipe emissions factor in? The general directive to reduce wait times does not dictate what those tradeoffs should be. Indeed, some choices may not even have occurred to policymakers, but surface only when the engineers come to design the algorithms, and are left to resolve the tradeoffs.

A growing literature identifies the social, political, and ethical dimensions of algorithms.[46] We address specific contextualized problems in Part II. For now, it is enough to highlight by way of example one especially important manifestation of an algorithm's politics: how a classification algorithm deals with false positives and false negatives. Take an algorithm that classifies objects in a train station as suspicious or not. Programmers must formalize the balance between the risk of false alarms and the risk of missing a dangerous object. In statistics, false positives are commonly known as "Type I Errors," and false negatives are known as "Type II Errors." Programmers must "tune" the algorithm to favor one kind of error over the other, or to treat them the same.[47] Nick Diakopoulos observes that algorithmic tuning "can privilege different

---

[44] Mittelstadt, et al., *supra* note 16, at 1 ("Operational parameters are specified by developers and configured by users with desired outcomes in mind that privilege some values and interests over others.").

[45] Harry Surden, *Values Embedded in Legal Artificial Intelligence* at 2 (2017), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2932333.

[46] *See* Nicholas Diakopoulos, Algorithmic Accountability, *Digital Journalism*, 3:3, 398-415, 400 (2015) (discussing the value choices embedded in data prioritization, classification, association, and filtering). *See also* Executive Office of the President National Science and Technology Council Committee on Technology, Preparing for the Future of Artificial Intelligence *(2016)*, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf; Felicitas Kraemer et al., Is There an Ethics of Algorithms?, 13(3) *Ethics and Information Technology 251–60 (2011)*; Bryce Goodman & Seth Flaxman, EU Regulations on Algorithmic Decision-Making and a " Right to Explanation"*(2016)*, arXiv:1601.08813 [stat.ML], https://arxiv.org/pdf/1606.08813.pdf.

[47] *See* Daniel Neyland and Norma Mollers, Algorithmic IF…THEN rules and the Conditions and Consequences of Power, *Information, Communication & Society*, 20:1, 45-62 (2016) (discussing algorithms that make just this kind of classification).

stakeholders in a decision, implying an essential value judgment by the designer of such an algorithm in terms of how false positive and false negative errors are balanced."[48]

For many algorithms, this tuning is not revealed. It was for Philadelphia's Adult Probation and Parole Department's risk prediction algorithm for violent recidivism among probationers. The tool 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.[49] Criminologist and statistician Richard Berk, who consulted on the program, estimates that between 29 percent and 38 percent of predictions end up being wrong – an error rate justified by a policy that it "much more dangerous to release Darth Vader than it is to incarcerate Luke Skywalker."[50] It turned out, however, that overclassifying probationers as high risk was problematic because they received more expensive services designed to smooth re-entry. The city went back to Berk and asked him to recalibrate the algorithm to reduce the size of the high-risk category. According to another project participant, the model was intentionally made less accurate "to make sure it produces the right kind of error when it does."[51]

The choice to privilege one type of error over another is one of dozens or hundreds of decisions that will inform the construction of a predictive algorithm. Some of these will be trivial and some consequential. Some will implement publicly stated policy objectives while others will have been left to programmers without policy direction. Cary Coglianese and David Lehr recognize that "[f]or agencies not accustomed to making moral valuations through any kind of formal process, let alone one that assigns them numbers, machine-learning algorithms will necessitate addressing questions of organizational and democratic decision making."[52]

---

[48] Diakopolous, *supra* note 46 at 401. *See also* Matthias Spielkamp, *Inspecting Algorithms for Bias*, MIT Technology Review (June 12, 2017), https://www.technologyreview.com/s/607955/inspecting-algorithms-for-bias/ (a sentencing algorithm can treat disparate groups "fairly" with respect to true positives (recidivism), but not with respect to false negatives (predicted recidivism that does not occur).

[49] Nancy Ritter, Predicting Recidivism Risk: New Tool in Philadelphia Shows Great Promise, National Institute of Justice No. 271 (2013), https://www.nij.gov/journals/271/pages/predicting-recidivism.aspx.

[50] Joshua Brustein, *This Guy Trains Computers to Find Future Criminals*, Bloomberg Technology (July 18, 2016), http://www.bloomberg.com/features/2016-richard-berk-future-crime/. See generally, *See* RICHARD A. BERK, STATISTICAL LEARNING FROM A REGRESSION PERSPECTIVE 13, 139-45 (2008) (discussion of scoring of different kinds of errors in machine learning algorithms).

[51] *Id.* (quoting Geoffrey Barnes)

[52] Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 Geo. L.J. 1147, 1218 (2017).

2. *Does the Algorithm Perform?*

Whatever the hidden policy choices an algorithm encodes, a government will presumably have a high-level explicit policy objective for a predictive algorithm, whether it is to reduce traffic wait time or to minimize recidivism among parolees. The public should be able to assess algorithmic performance in achieving the stated goals. This is a relatively simple question of utility as assessed by statistical performance in fitting the data to the desired outcome. Even here, of course, there are a variety of measures of performance, and it is important to understand what each measure represents. For example, one popular measure used for predictive algorithms is area under the receiver operating characteristic (ROC) curve. In a single number between 0.5 and 1, it provides an assessment of how much better an algorithm is than a random assignment of cases at avoiding both false positives and false negatives. However, it has some limitations – it can only be applied when the output of the algorithm is a score that ranks subjects from least to most likely to be associated with some outcome – and it provides only one perspective on the relative success of the algorithm. Other measures may focus on other aspects of performance. For example, "goodness of fit" tests may reveal that although a model is quite good overall at predicting the risk of a particular outcome, its predictions that subjects are among the riskiest 10% are significantly accurate than its predictions that subjects are among the least risky 10%.[53] A variety of additional measures are available, and the area of predictive algorithm assessment continues to develop.[54]

There are many reasons an algorithm could be ineffective.[55] It could be trained on bad data inputs (garbage in, garbage out).[56] Errors may also result from faulty inductive

---

[53] *See, e.g.*, Alberto Maydeu-Olivares & C. Garcia-Forero, Goodness-of-fit testing, in *International encyclopedia of education* 190 (2010).

[54] *See, e.g.*, Ewout W. Steyerberg, Andrew J. Vickers, Nancy R. Cook, Thomas Gerds, Mithat Gonen, Nancy Obuchowski, Michael J. Pencina, & Michael W. Kattan, Assessing the performance of prediction models: a framework for some traditional and novel measures, 21 Epidemiology 128 (2010); Mauno Vihinen, How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis, 13 BMC Genomics S2 (2012), https://doi.org/10.1186/1471-2164-13-S4-S2; Dean Abbott, Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst 283-304 (2014); Scott Fortmann-Roe, Accurately Measuring Model Prediction Error, http://scott.fortmann-roe.com/docs/MeasuringError.html

[55] *See* Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights, Executive Office of the President (May 2016) (discussing poorly selected data; incomplete, incorrect or outdated data; selection bias; unintentional perpetuation and promotion of historical biases), https://obamawhitehouse.archives.gov/blog/2016/05/04/big-risks-big-opportunities-intersection-big-data-and-civil-rights.

[56] Wikipedia, Garbage In, Garbage Out, http://en.wikipedia.org/wiki/Garbage_in,_garbage_out (a computer science term expressing the informal rule that the quality of a computer's output is only as good as the quality of its input.

reasoning, data selection, and factor weighting.[57] Another point of failure in the broader algorithmic process may be at the implementation phase.[58] Unless the algorithmic prediction is self-executing, human beings have to understand the prediction in order to decide how much weight to give it. In the municipal context, government workers will often be responsible for selecting and inputting data as well. While validation studies can help to ensure that an algorithm is achieving the desired goal, cash-strapped governments may not require validation studies before or after implementation, or they may not be conducted properly. The results of validation studies, as well as information about their design, should all be subjected to public scrutiny.

3. Is the Algorithm Fair?

An algorithm may perform well in terms of achieving desired outcomes, but come up short on equitable measures. There is a strong public interest in ensuring that predictive algorithms are designed and executed justly, especially when they impact individuals. Fairness concerns will generally matter much less to developers than the performance of an algorithm, and may not figure in an engineer's remit at all.[59]

Government use of predictive algorithms poses an inherent challenge to traditional notions of fairness.[60] By their nature, predictive models are simplifications, which do not take into account all possible relevant facts about subjects, and they therefore treat people as members of groups, not as individuals.[61] Generalizations are inherent in this process. For sensitive decisions, particularly where individual liberty is at stake, decisionmakers like judges and social workers are expected to exercise human judgment over algorithmic predictions. In theory, the algorithmic edict is advisory only.

---

[57] [cite to sources of algorithmic failure – AI literature]. *See also* Houston Federation of Teachers v. Houston Independent School District, Civil Action H-14-1189, Amended Summary Judgment Opinion, at 13 (S.D. Tx. May 4, 2017) (noting that an algorithmic score "might be erroneously calculated for any number of reasons ranging from data-entry mistakes to glitches in the computer code itself. Algorithms are human creations, and subject to error like any other human endeavor.").

[58] [cite data and society ethnographic study]

[59] Nick Seaver, Knowing Algorithms, *Media in Transition* 8 at 2 (2013), http://nickseaver.net/s/seaverMiT8.pdf (the policy implications of an algorithm are "strictly out of frame" for algorithm developers). This is a challenge computer science is beginning to explore. *See, e.g.,* Michael Feldman et al, Certifying and Removing Disparate Impact (2016) (presenting a test for disparate impact in algorithmic processes and a method by which data might be made unbiased); Sorelle A. Friedler, Carlos Scheidigger & Suresh Venkatasubramanian., On the (im)possibility of fairness (submitted 2016), https://arxiv.org/pdf/1609.07236.pdf..

[60] Fairness itself is subject to different definitions; the definition selected will affect assessments of algorithmic fairness. *See* Friedler et al., *supra* note 60 (recommending that computer scientists make more explicit what notion of fairness they seek to represent in algorithms).

[61] *See, e.g.*, O'Neil, *supra* note 5, at 20-23; Mittelstadt et al., *supra* note 16, at 8.

In practice, decisionmakers place heavy reliance on the numbers, raising the stakes for their fairness.[62]

The most discussed algorithmic fairness question has been whether predictive algorithms are likely to introduce or perpetuate invidious discrimination on the basis of race, gender, or another protected characteristic.[63] There are additional forms of discrimination that are of concern, such as whether an algorithm incidentally disfavors (and therefore, disincentivizes) certain behaviors. For example, if use of the mental health system correlates with increased risk of child endangerment, then an algorithm trained on this data might include mental health system use as a factor in its risk assessment. Use of the mental health system may or may not correlate with membership in a protected class. But an algorithm that penalizes those who seek mental health treatment raises fairness concerns, as well as larger welfare concerns if those who would be aided by mental health treatment choose not to seek it to avoid child welfare interventions.

Fairness and performance may sometimes be correlated. A classic example is the early Google facial recognition algorithm. It was trained on the faces familiar to the engineers who built it, which were mostly white.[64] As a result, the program classified white-skinned human faces as human, but often classified dark-skinned human faces as animal. Retraining the algorithm using human faces of all skin colors would make it perform better overall, as well as reduce the disparity of inaccuracies between light- and dark-skinned human faces. When making an algorithm fairer would actually increase its utility, we can expect that rigorous analysis of performance will also lead to greater fairness.

In some cases, however, there may be a trade-off between fairness and performance. Inclusion of an individual's group membership may enhance algorithmic utility if the observed correlations are not simply duplicative of other correlations in the data. Take the correlation that some data analysis has found between convicted felons from certain zip codes and higher rates of recidivism.[65] That correlation might not increase

---

[62] [cite Loomis; Allegheny County evidence. [John Danaher, *The Threat of Algocracy: Reality, Resistance and Accommodation*, 29 PHIL. & TECH. 245 (2016)].

[63] *See, e.g.*, O'Neil, *supra* note 5; Joh, supra note 3; Barocas & Selbst, *supra* note 6; Pauline T. Kim, *Data-Driven Discrimination at Work*, __ WM. & MARY L. REV. __ , (forthcoming 2017), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2801251; Andrew D. Selbst, *Disparate Impact in Big Data Policing* __ (forthcoming 2017), *available at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2819182.

[64] Clare Garvie & Jonathan Frankle, Facial-Recognition Software Might Have a Racial Bias Problem, *The Atlantic* (Apr. 7 2016), Anupam Chander, T*he Racist Algorithm*? 115 MICH. L. REV. 1023 (2017) (book review).

[65] *See* Angwin et al., *supra* note 6. *But see* COMPAS *Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*, http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf, and Anthony W. Flores, et al, *False*

the predictive power of an algorithm if the algorithm were also using, say, employment history as a factor. Zip codes and employment history might turn out to be nearly covariants, but with employment history as the better predictor. Why that could be the case would be a further question, but one theory might be that zip codes were only serving as a weak signal of future employment, due to geographic clustering of unemployment, and were therefore not improving on predictions that directly factored in employment history. Conversely, however, inclusion of zip code information might demonstrably increase the predictive power of an algorithm, pointing to some correlation that was not covered by any other included variable or characteristic.

Even if use of zip codes improved an algorithm's predictive power, however, a public agency may decide to exclude them. Due to residential segregation, zip codes may turn out to be close proxies for race. A public agency may decide that race should not be taken into account even if it has predictive power. It may conclude that it is extremely unlikely that skin color itself has any causal relation to the desirable or undesirable behavior, and that using race as a shortcut for whatever might actually have some causal relation would perpetuate "a history of purposeful unequal treatment"[66] based on "an immutable characteristic determined solely by the accident of birth."[67] In other words, if some marginal increase in accuracy is almost certainly accompanied by an increase in unfairness to a protected class, a public agency may choose fairness over accuracy. Of course, in some situations, taking race into account may simply reinforce historical patterns of bias. Minority neighborhoods historically subject to more intensive policing will have higher arrest and re-arrest rates, and then be recommended by the algorithm for more policing, and so on.[68] A historical pattern of discriminatory treatment will thus *cause* higher observed crime rates in the zip codes that the algorithm predicts are at higher risk for criminal activity.

Jurisdictions have dealt with such fairness concerns in different ways. The Oakland Police Department decided not to use a predictive algorithm (PredPol) at all, having concluded that "officers would have been deployed to mostly lower-income minority neighborhoods where the previous drug crimes were recorded."[69] By contrast, cities like Philadelphia and Chicago are using predictive policing programs, but their vendor (Azavea, Inc., developers of HunchLab) has decided to deemphasize some arrest data,

---

Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks," http://www.crj.org/page/-/publications/rejoinder7.11.pdf. *See also* Rhema Vaithianathan, et al., Developing Predictive Models to Support Child Maltreatment Hotline Screening Decisions: Allegheny County Methodology and Implementation. Aukland University of Technology, New Zealand: Centre for Social Dynamics. September 2016 (discussing zip codes and other proxies for race).

[66] San Antonio Ind. School Dist. v. Rodriguez, 411 U.S. 1, 28 (1973).

[67] Frontiero v. Richardson, 411 U.S. 677, 686 (1973).

[68] *See* O'Neil, *supra* note 5, at xx.

[69] Emily Thomas, *Why Oakland Police Turned Down Predictive Policing* (12/28/16), http://motherboard.vice.com/read/minority-retort-why-oakland-police-turned-down-predictive-policing

particularly data concerning drug-related and nuisance crimes, in creating its policing models, to avoid likely systemic bias.[70] Ultimately, unless a predictive algorithm is rendered sufficiently transparent, we won't know whether automated decisionmaking and risk prediction accords with our substantive commitments to fairness.[71]

4. Does the Algorithm Enhance or Diminish Governmental Capacity?

We have noted that predictive algorithms promise to help governments make better decisions by replacing hunches with more objective correlations. We have also noted that they may inject systemic bias or error into decisionmaking. There is a further danger that algorithmic governance, impervious to critical evaluation while also displacing human decisionmaking, will hollow out the decisionmaking capacity of public servants. Contributing factors could include unwarranted deference to the algorithm, insufficient understanding of algorithmic processes, and/or atrophied competence to use human judgment.

Government officials may defer to algorithmic output even when it is erroneous, discriminatory, or framed in terms of categories that are too coarse or outcomes that are too narrow. When the "machine says so," it can be difficult for rushed and over-extended human decisionmakers to resist the edict. As Harry Surden notes, judges may "give more deference to computer-based recommendations, compared to similar human based recommendations, given the aura of mechanistic legal neutrality that recommendations made by computers analyzing data may seem to have."[72] According to Michael Ananny, "algorithmic categories . . . signal certainty, discourage alternative explorations, and create coherence among disparate objects," thereby reifying the algorithmic model's choices.[73]

Second, when algorithmic output is uninterpretable – when the decision path is not explained - government officials have no way of knowing whether and how the factors they're facing accord with the factors that produced the algorithmic recommendation. Suppose that the criminal defendant the algorithm is scoring has been blinded or has had a child. Might those facts justify deviating from the algorithm's risk prediction, or are they accounted for? If the algorithm is opaque, the government official cannot know how to integrate its reasoning with her own, and must either disregard it, or

---

[70] *See* A Citizen's Guide to HunchLab 26 (July 11, 2017 draft), http://robertbrauneis.net/algorithms/HunchLabACitizensGuide.pdf. Azavea also recommends introducing a small degree of randomness into the algorithm to make a "probabilistic selection of locations," for patrol, in part to counter bias. *See id.* at 10-11.

[71] See https://joanna-bryson.blogspot.co.uk/2017/07/three-very-different-sources-of-bias-in.html ("The way to deal with [bias] is to insist on the right to explanation, on due process. All algorithms that affect people's lives should be subject to audit").

[72] *See* Surden, *supra* note 45, at 2. *See also* danah boyd and Kate Crawford, Critical Questions for Big Data, *Inf. Commun. Soc.* 15(5):662–679 (2012) (identifying mistaken belief in objectivity as one of the pitfalls of reliance on big data analytics).

[73] Ananny, *supra* note 5 at 103. *See also* Barocas & Selbst, *supra* note 6.

follow it blindly. Thus, as Christopher Church and Amanda Fairchild have said, "The reasoning behind an algorithm's prediction is critically important. The algorithm must not only be able to accurately identify the high risk cases . . . but must also be able to provide contextual reasoning for why certain cases are being flagged."[74]

Third, over time, deference to algorithms may weaken the decisionmaking capacity of government officials along with their sense of engagement and agency. The "de-skilling" of human beings through automation has become a widely-studied phenomenon,[75] and it will undoubtedly spread to public administration. Ethicists have also examined how computer systems can undermine peoples' sense of their own moral agency. When "human users are placed largely in mechanical roles, either mentally or physically," and "have little understanding of the larger purpose or meaning of their actions . . . human dignity is eroded and individuals may consider themselves to be largely unaccountable for the consequences of their computer use."[76] The same can be said more specifically about predictive algorithms and the government officials who use them. For example, police personnel who are instructed by algorithm exactly where and how to patrol may lose their own awareness of crime risks, and be unable to responsibly deviate from the algorithm's instructions.[77]

The decision path an algorithmic process took to generate a recommendation should therefore ideally be disclosed to the government officials tasked with implementation. That disclosure would help government officials to feel responsibility for the decisions they make, and to cultivate skills appropriate to decisionmaking in their fields. The public should be able to find out whether government officials have been trained in the logic and limitations of the algorithms they use, so that citizens can assess whether the algorithm may be eroding the skills, agency, and accountability of public officials.

D. Transparency

It will be possible to assess a predictive algorithm's politics, performance, fairness, and relationship to governance only with significant transparency about how the algorithm

---

[74] Christopher E. Church and Amanda J. Fairchild, *In Search of a Silver Bullet: Child Welfare's Embrace of Predictive Analytics*, 68 Juvenile and Family Ct. J. 67, 78 (2017).

[75] *See, e.g.*, Nicholas Carr, The Glass Cage: How Our Computers are Changing Us (2014).

[76] Batya Friedman & Peter H. Kahn, Jr., *Human Agency and Responsible Computing: Implications for Computer System Design*, 17 J. of Systems Software 7 (1992); *see* Helen Nissenbaum, *Accountability in a Computerized Society*, 2 Science & Engineering Ethics 25 (1996).

[77] On the "de-skilling" of police occasioned by computerized risk analysis, *see* Richard V. Ericson & Kevin D. Haggerty, Policing the Risk Society 447 (1997). On the dangers of "de-skilling, the erosion of professional discretion, and . . . a process of de-professionalization" stemming from use of algorithms by probation decisionmakers, see Gwen Robinson, *Implementing OASys: Lessons from Research into LSI-R and ACE*, 50 Probation J. 30, 33 (2003); Diana Wendy M. Fitzgibbon, *Risk Analysis and the New Practitioner: Myth or Reality?*, 9 Punishment & Society 87, 90 (2007)

works. Algorithmic opacity is a problem widely recognized and variously defined.[78] As a general matter[79] and with respect to public sector applications, commentators recognize the need for more transparency in the implementation of predictive algorithms.[80] So do courts presented with cases of first impression about the due process rights of individuals affected by algorithmic judgment to know the reasons why the machine "said so."[81]

To be sure, there has always been risk of inefficacious or biased decisionmaking by government agents. We cannot know if a judge deciding on pre-trial flight risk is properly considering risk factors. Why should automated reasoning be revealed to us when human reasoning was not? First, more transparency is better than less when it comes to decisions to use government force, deprive citizens of their liberty, or allocate public resources. The formalization of predictions in an algorithm may give us the opportunity to test whether those predictions are inaccurate or unfair, and properly viewed, that is part of the promise of algorithms.

Second, predictive algorithms pose new risks of unfairness and error even if they improve overall decision making. This is because where a problem exists, it will be worse and more durable. Predictive algorithms are typically used to guide decisions throughout a governmental unit – all criminal judges in a jurisdiction, for example -

---

[78] For an exploration and taxonomy of various kinds of algorithmic opacity, see Andrew D. Selbst & Salon Barocas, *Regulating Inscrutable Systems*, *available at* http://www.werobot2017.com/wp-content/uploads/2017/03/Selbst-and-Barocas-Regulating-Inscrutable-Systems-1.pdf.

[79] *See, e.g.*, Frank Pasquale, *Restoring Transparency to Automated Authority*, 9 J. TELECOMM. & HIGH TECH. L. 235 (2011); Ed Felten, *Accountable Algorithms*, FREEDOM TO TINKER (Sept. 12, 2012), https://freedom-totinker.com/2012/09/12/accountable-algorithms/; Commissioner Julie Brill, Federal Trade Commission, Scalable Approaches to Transparency and Accountability in Decisionmaking Algorithms: Remarks at the NYU Conference on Algorithms and Accountability (2015), https://www.ftc.gov/system/files/documents/public_statements/629681/150228nyualgorithms.pdf; Nicholas Diakopolous, Revealing Algorithms, https://epic.org/algorithmic-transparency/. *But see* Mike Ananny & Kate Crawford, Seeing Without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability, New Media & Society(2016), *available at* http://journals.sagepub.com/doi/abs/10.1177/1461444816676645 (arguing that some algorithmic processes may be inherently non-transparent and, therefore, should be rendered accountable in other ways).

[80] Lee P. Breckenridge, *Water Management for Smart Cities: Implications of Advances in Real-Time Sensing, Information Processing, and Algorithmic Controls*, 7 G.W. J. OF ENERGY & ENV. L. 153, 162 (2016) (identifying in the context of smart city water management the dangers of "automated processes for sensing, analyzing, and responding to complex information… unless the administrative processes for adopting these systems are themselves made accessible, transparent, and subject to ongoing and meaningful review.").

[81] *See, e.g.*, Houston Federation of Teachers v. Houston Independent School District, *supra* note 57 (allowing teachers' due process action against public school district for implementing a teacher evaluation algorithm that is impervious to investigation).

and even across many local and state governments. [82] This is the problem that Cathy O'Neil identifies as the scalability of algorithms.[83] The ability of these algorithmic processes to scale, and therefore to influence decisions uniformly and comprehensively, magnifies any error or bias that they embody, and increases the importance of rendering them transparent.

The challenge is to specify a degree and form of transparency that is meaningful for the public and practical for developers and governments. Parts II and IV below identify the kind of information that should be revealed about publicly deployed algorithms. Here, we unpack several layers of transparency, and highlight the centrality of transparency to open records laws.

Algorithmic processes can be opaque and resistant to knowing in different ways. Following Frank Pasquale, commentators have focused on concealment of algorithmic formulas, inputs, and rules of procedure in a "black box."[84] Disclosing the algorithm's formal components might reveal mistakes in the algorithm itself – it might reveal, for example, that the algorithm sometimes generates results outside of the range to which it is supposed to be limited, or conversely that its results will always be more limited than the range it is supposed to produce.

Algorithms should be capable of disclosure in some combination of mathematical and logical notation and natural language.[85] To be implemented by computer, they must be coded in a programming language. Disclosure of the computer code may be appropriate if there is a concern that the computer implementation may be incorrect.[86] However, the computer program will usually be significantly harder for human beings to read and understand than mathematical or logical notation or natural language, and hence disclosure of computer code may be the less helpful alternative.

---

[82] For example, the Arnold Foundation's PSA-Court is used by three entire states – Arizona, Kentucky, and New Jersey – and in 35 other jurisdictions. *See* http://www.arnoldfoundation.org/initiative/criminal-justice/crime-prevention/public-safety-assessment/.

[83] *See* O'Neil, *supra* note 5, at 29-31.

[84] *See* Pasquale, *supra* note 5, at 1-18; *see also* O'Neil, *supra* note 5, at 28-31, Selbst & Barocas, *supra* note 78, at 9.

[85] For one example of such disclosure, see Marie VanNostrand & Gina Keebler, Pretrial Risk Assessment in the Federal Court 48 (2008), http://tiny.cc/r3qrmy (disclosing in mathematical notation a formula to predict risk of failure to appear at trial and risk of crime upon pretrial release, and in natural language a description of each factor used).

[86] *See* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1268 (2008) (stating that from September 2004 to April 2007, code writers embedded over 900 incorrect rules into Colorado's public benefits system, which resulted in hundreds of thousands of incorrect eligibility determinations and benefits calculations for Medicaid and food stamps during this period); *id.* at 1270 (stating that code writers incorrectly translated policy into California's automated public benefits system, causing over- and underpayments and improper terminations of public benefits).

Even if the algorithm's formal components are revealed, the algorithmic process may still not be capable of evaluation. The algorithm's claim of validity is not limited to compliance with the algorithm's own rules. The claim rests on correlations between facts and outcomes in an underlying dataset. We cannot interrogate this claim without knowing something about the training data. How was the data selected, why were particular rules of operation chosen while others were rejected, and what steps were taken to validate those choices?[87] Access to the underlying data or at least descriptions of it would help us understand how strong the purported correlations actually are, what the sample size was, and other matters that affect statistical validity.

Another type of information now typically sunk in obscurity is the public purpose for which the algorithm was developed, the contract terms that govern data ownership and access, and plans for validation and follow-up. Sometimes, this information will also either explicitly or implicitly address some of the policy tradeoffs the algorithm entails. All of this will be important to assess whether the algorithm is effective, fair, and otherwise politically acceptable.

We acknowledge that even if all the information identified above were revealed, it might still be impossible to understand the results of an algorithmic process. This is because transparency does not necessarily render an algorithm "interpretable."[88] If an algorithm uses hundreds of unweighted inputs in a complex decision tree in which a single input might appear at multiple junctures, we can't necessarily figure out which inputs were decisive in a particular case.[89] This makes it particularly difficult to understand whether the algorithm correlates with our sense of fairness, and it makes it difficult for government officials to assess algorithmic output in light of their own sense of a situation, requiring them either to ignore that output, or ignore their own judgment, and perhaps eventually to lose that judgment.

---

[87] In a case of first impression for the Fifth Circuit, Houston teachers have sought "the equations, computer source codes, decision rules, and assumptions" built into a privately-created evaluation algorithm used by the school district. Houston Teachers Federation, *supra* note 57, . at 17.

[88] Of course, people can sometimes provide us with explanations of their reasoning process. However, we have no guarantee that these explanations actually match how they came to their decisions. *See* Zachary C. Lipton, The Mythos of Model Interpretability 98 (2016), http://arxiv.org/abs/1606.03490v2 (noting this, and defining the ability of a model to be explained after the fact as "post hoc interpretability").

[89] If a model tries to predict parolee recidivism by assigning weights to a few factors like prior violent offenses and age, we can understand and explain what it's doing. Suppose, however, that the model uses over 1000 factors to predict parolee recidivism, some of which don't seem to have any intuitive causal connection to recidivism (say, height). Suppose moreover, that the model uses a complex decision tree featuring many factors multiple times. It will be difficult to understand how influential each factor is in a particular case or over a range of cases, or to articulate the model's theory of causation (if any).

Lastly, algorithmic processes may be dynamic, their rules changing constantly to fit new patterns in the data.[90] As a result, the code and data sets that are released to the public at T1 – even if "interpretable" in isolation – may bear little resemblance to the process that is conducted at T2. Dynamic algorithms are, as Rob Kitchin says, "ontogenetic in nature," subject to being "edited, revised, deleted and restarted."[91] We will not be addressing this kind of dynamism in large part because the local and state government actors we studied are not yet using these constantly-adjusted predictive algorithms.

Just as transparency does not necessarily support interpretability, transparency is not coextensive with accountability.[92] It is merely a means. An algorithmic process is *perfectly* transparent when its rules of operation and process of creation and validation are completely known. *Meaningful*, sufficing transparency is a lower standard. An algorithmic process is accountable when its stakeholders, possessed of meaningful transparency, can intervene to effect change in the algorithm, or in its use or implementation.[93] Algorithmic accountability in the public sphere requires that government actually be held accountable for the algorithms it deploys. For this to happen in practice, it could well require public education and political processes that are beyond what we can address here. But meaningful transparency will be the necessary first step.

II. Results of Open Records Requests for Algorithms

Open data practices are probably the best way to make transparent aspects of the algorithms used in government.[94] That is, governments should reveal the relevant structures, logic and policies of the algorithms voluntarily from the outset. Amendments to the Federal FOIA in 2016 codified this preference for a "push" method of transparency, which reduces the load on "pull" requests for government records.[95] Yet governments have not been pushing out information about the algorithms they use. There is a big gap between the importance of algorithmic processes for governance and

[90] See Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights, Executive Office of the President (May 2016) at 8 (As machine learning methods advance "it may become more difficult to explain or account for the decisions machines make through this process unless mechanisms are built into their designs to ensure accountability.").

[91] Kitchin, Thinking Critically about and Researching Algorithms, *supra* note 5, at 18.

[92] Kroll *et al.*, *supra* note 25, at 657-660. *See also* Ananny, *supra* note 5, at 109.

[93] *See* Kroll et al., *supra* note 25, at 657-660; Selbst & Barocas, *supra* note 78, at 15.

[94] *See generally*, Joshua Tauberer, The Principles and Practices of Open Government Data 10 (2nd ed. 2014).

[95] The *FOIA Improvement Act of 2016* amended 44 U.S.C. § 3102 (agencies must establish "procedures for identifying records of general interest or use to the public that are appropriate for public disclosure, and for posting such records in a publicly accessible electronic format.").

public access to them.[96] In the absence of push transparency, open records requests are the next best way to close the gap and vindicate the public's interest in understanding the algorithms that are being applied to them and their fellow citizens. We tested how responsive governments are to such requests for information concerning predictive algorithms and associated data analytics. After introducing open records laws in general, we discuss our results.

A. Open Records Laws

The state freedom of information laws we relied on in seeking information about algorithmic processes all have the same central purpose: to reveal the workings of government to the people.[97]

Freedom of information laws have as their principal goal accountable government. In signing the original FOIA in 1964, President Johnson expressed "a deep sense of pride that the United States is an open society in which the people's right to know is cherished and guarded."[98] With FOIA amendments establishing deadlines for agency responses a decade later, Congress celebrated "[o]pen government . . . as the best insurance that government is being conducted in the public interest."[99] And when Congress affirmed in 1996 that the central transparency mandate of FOIA applied to electronic records, the Senate Committee Report explained that government transparency "is consistent with our democratic form of government by furthering the interests of citizens in knowing what their Government is doing."[100] The courts have consistently held that FOIA embodies "a general philosophy of full agency disclosure."[101]

---

[96] *See* Diakopoulos, Algorithmic Accountability Reporting: On the Investigation of Black Boxes 2 (2013), *available at* https://academiccommons.columbia.edu/catalog/ac:2ngf1vhhn4 ("What we generally lack as a public is clarity about how algorithms exercise their power over us.").

[97] *See, e.g.*, New York Freedom of Information Law, N.Y. Pub. Off. Law § 86(4) ("The people's right to know the process of governmental decision-making and to review the documents and statistics leading to determinations is basic to our society. Access to such information should not be thwarted by shrouding it with the cloak of secrecy or confidentiality. The legislature therefore declares that government is the public's business and that the public, individually and collectively and represented by a free press, should have access to the records of government.").

[98] Statement by President Lyndon B. Johnson upon Signing Pub. L. 89-487 on July 4, 1966, in Attorney General's Memorandum on the Public Information Section of the Administrative Procedure Act (1967), available at http://www.justice.gov/oip/67agmemo.htm.

[99] S.REP. NO. 93-854, at 1 (1974).

[100] S. REP. NO. 104-272, at 5 (1996), http://nsarchive.gwu.edu/nsa/foialeghistory/104_cong_reports_efoia_senate.pdf. *See generally* Michael Schudson, Rise of the Right to Know (2015).

[101] Department of Air Force v. Rose, 425 U.S. 352, 360 (1976); *see also* N.L.R.B. v. Robbins Tire & Rubber Co., 437 U.S. 214, 236 (1978).

Animated by the same transparency principles, the open records statutes of all 50 states and the District of Columbia provide individuals with the right to access government records, subject to various exemptions. These include exemptions to protect individual privacy, criminal investigatory material, and agency deliberative processes. [102] Almost all the laws also exempt trade secrets, which we discuss below. FOIA applies to "agency records" – an undefined term.[103] The Supreme Court understands "agency records" to include any records that an agency 1) creates or obtains and 2) has control of at the time of the FOIA request.[104] Although state laws more typically use the term "government records," the coverage is similar.[105]

FOIA covers digital records, including software and databases.[106] Some state laws expressly include software as a public record.[107] Under New Jersey's open records statute, for example, a "government record" includes any "data processed or image processed document" and "information stored or maintained electronically" if it has been made, maintained, or received by a State officer or employee in the course "of his or its official business."[108] Other state statutes expressly *exclude* software from public

---

[102] *See, e.g.*, 5 U.S.C. § 552(b)(1)–(9). With respect to FOIA, the Supreme Court "has repeatedly stated that these exemptions from disclosure must be construed narrowly, in such a way as to provide maximum access." Vaughn v. Rosen, 484 F.2d 820, 823 (D.C. Cir. 1973).

[103] 5 U.S.C. § 552(a)(4)(B)

[104] *See* Forsham v. Harris, 445 U.S. 169, 182 (1980).

[105] *See, e.g.*, ARIZ. REV. STAT. § 41-151.18; CAL. GOV'T CODE § 6252.

[106] *See, e.g.*, 5 U.S.C.§(f)(2)(A)("'record' and any other term used in this section in reference to information includes any information that would be an agency record subject to the requirements of this section when maintained by an agency in any format, including an electronic format."). Some state open records exclude certain kinds of software. *See, e.g.*, California Open Records Law (excluding computer software "developed by a state or local agency … includ[ing] computer mapping systems, computer programs, and computer graphic systems." (§§ 6254.9(a),(b))).

[107] *See generally*, Cristina Abello, Reporters Committee for Freedom of the Press, Access to Electronic Communications (2009), *available at*http://www.rcfp.org/rcfp/orders/docs/ELECCOMM.pdf; Andrea G. Nadel, Annotation, *What Are "Records" of Agency Which Must Be Made Available under State Freedom of Information Act*, 27 A.L.R.4th 680 (Supp. 2014); Marjorie A. Shields, Annotation, *Disclosure of Electronic Data under State Public Records and Freedom of Information Acts*, 54 A.L.R.6th 653 (Supp. 2014).

[108] N.J. Stat. Ann. § 47:1A-1.1 (West 2015). [NJ Supreme Court decision on coverage of "information."] *Cf.* N.Y. Pub. Off. Law § 86 (4) (McKinney 2003) (defining "record" as "any information kept, held, filed, produced or reproduced by, with or for an agency or the state legislature, in any physical form whatsoever"); Fla. Stat. Ann. § 119.011 (12) (West 2016) (defining "public records" to include "data processing software"); 5 Ill. Comp. Stat. Ann. 140/2 § 2 (c) (2016) ("records" includes "electronic data processing records.").

records.[109] Still others have not addressed the issue.[110] Whether or not a member of the public is rightfully able to insist on the production of software under a state open records act will rarely be the most important issue for algorithmic transparency given that meaningful transparency can be achieved through other kinds of records.

The most formidable obstacle will be ownership of the record. Most open records laws only cover government records. To the extent that private contractors have exclusive control of records, those records may be beyond the reach of transparency laws. FOIA provides that when records are "maintained for an agency by an entity under Government contract, for the purposes of records management," those records remain "agency records" subject to FOIA disclosure. This covers situations where an agency contracts with a private vendor to maintain records, such as police camera video.[111] These records are agency records even though they reside on private servers. However, where a private party generates records for its *own* purposes and never deposits them with an agency, such records are likely to fall outside FOIA and state open records acts.[112] In the case of algorithms, these may include the training data and documentation of the process of constructing and validating the algorithm. As discussed below, public access to these records will depend on the insistence of government agencies on data ownership and/or possession of records.

B. Requests, Responses, Conclusions

We filed open records requests covering six different programs featuring predictive algorithms. In some cases, we also engaged in direct communication with the contractors that developed the algorithms. This section discusses those requests and communications, and the government responses. The six programs are: Public Safety Assessment; Eckerd Rapid Safety Feedback; Allegheny Family Screening Tool; PredPol; HunchLab; and New York City Value-Added Measures. Generally, we found wide variation in whether jurisdictions responded; whether they claimed an open records exemption; and if not, what information they provided. However, only one of the jurisdictions, Allegheny County, was able to furnish both the actual predictive algorithms it used (including a complete list of factors and the weight each factor is given) and substantial detail about how they was developed. Some developers were

---

[109] *See, e.g.*, AS 40.25.220(3) (definition of "public records" does not include "proprietary software programs.")

[110] *See, e.g.*, Ala. Code § 36-12-40 (providing that "[e]very citizen has a right to inspect and take a copy of any public writing of this state, except as otherwise expressly provided by statute" without addressing meaning of "writing.")

[111] *See* Alexandra Mateescu, et al., Police Body-Worn Cameras, Working Paper (2 Feb. 2015) (Data & Society Research Institute) at 9 (discussing police department storage of police body camera footage in third-party cloud servers), available at https://www.datasociety.net/pubs/dcr/PoliceBodyWornCameras.pdf.

[112] State treatment of records held by private entities is complex and varied. For a review, see Alexa Capeloto, *Transparency on Trial: A Legal Review of Public Information Access in the Face of Privatization*, 13 Conn. Pub. Int. L.J. 19, 27 (2013).

also more forthcoming than others. While the Arnold Foundation, developer of Public Safety Assessment, has disclosed its relatively simple algorithms to the public, it provided us next to nothing about its development process, while Azavea, Inc., developers of HunchLab, disclosed much more. These results suggest that transparency is a choice that jurisdictions and their vendors make – a choice having less to do with immutable trade secrets or confidentiality concerns than with a culture of disclosure.

1. Public Safety Assessment – Pretrial Release

Public Safety Assessment (PSA) is a pretrial risk assessment tool developed by the Laura and John Arnold Foundation, designed to assist judges in deciding whether to detain or release a defendant before trial.[113] As of this writing, it is being used in 38 jurisdictions, including the entire states of Arizona, New Jersey, and Kentucky.[114] PSA includes three different risk assessment algorithms, which are intended to assess the risks that a released defendant will, respectively, fail to appear for trial; commit a crime while on release; and commit a violent crime while on release.

The three algorithms operate by assigning points based on nine facts about the defendant's criminal history; some facts are used for only one or two of the algorithms, while others are used for all three. For the failure to appear and commission of crime assessments, the raw point scores are converted to a six-point scale, in which one is lowest risk and six is highest risk. For the commission of violent crime assessment, the raw score is converted into a binary yes/no answer; a crime committed is either likely to be violent, or likely not to be violent.[115]

Unlike some of the other algorithms, PSA is relatively simple – it can be implemented without a computer by tallying up points for various factors, and then applying a conversion formula to obtain the final risk assessment. The PSA algorithms, unlike many others, are fully disclosed. However, the Arnold Foundation has not revealed how it generated the algorithms, or whether it performed pre- or post-implementation validation tests, and if so, what the outcomes were. Nor has it disclosed, in quantitative or percentage terms, what "low risk" and "high risk" mean: is the chance that a "low risk" defendant will fail to appear one in ten or one in five hundred? Is the chance that a "high risk" defendant will fail to appear twice that of a low risk defendant or fifty times?

---

[113] *Public Safety Assessment*, LJAF, available at http://www.arnoldfoundation.org/initiative/criminal-justice/crime-prevention/public-safety-assessment/

[114] *See* http://www.arnoldfoundation.org/initiative/criminal-justice/crime-prevention/public-safety-assessment/.

[115] A description of all three algorithms, including factors, raw point allocations, and conversion from raw scores to final output, is available at http://www.arnoldfoundation.org/wp-content/uploads/PSA-Risk-Factors-and-Formula.pdf

To see whether the courts that were using PSA had answers to these or similar questions, we sent open records requests regarding the PSA program to 16 different courts. We sent a large number of requests – the largest for any of the algorithms we chose to study – in part because we knew that many open records laws exempt courts from most disclosures. Of the five courts that responded by providing some documents,[116] four of them – the Mesa Municipal Court and the Pima and Navajo County Court systems in Arizona, and the San Francisco Superior Court system in California – stated that they could not provide information about PSA because that information was owned and controlled by the Arnold Foundation.[117] Three of those four (Pima County, Navajo County, and San Francisco) sent us copies of their Memoranda of Understanding with the Arnold Foundation, which contained identical language prohibiting the courts from disclosing any information about the PSA program.[118]

---

[116] Four courts never responded, and one acknowledged receipt of our request, but did not further respond. Four courts responded that they had no relevant documents, and two courts rejected our requests, concluding that the relevant open records laws did not require them to provide the material we requested. For example, the Superior Court of New Jersey rejected our request, responding that its rules exempt from disclosure "records relating to the Pretrial Services Program" and "notes, memoranda or other working papers maintained in any form by or for the use of a justice, judge or judiciary staff member in the course of his or her official duties." Letter of December 22, 2016 from Michelle M. Smith, Clerk of the Superior Court, available at https://www.muckrock.com/foi/new-jersey-229/new-jersey-superior-court-public-safety-assessment-court-28835/#file-114392. The Allegheny County Court also rejected our request, on the ground that the Pennsylvania Right-to-Know Law applies to the judiciary only with respect to financial records, and our request was not for financial records. Email of October 6, 2016 from Christopher H. Connors, Chief Deputy Court Administrator, available at https://www.muckrock.com/foi/allegheny-county-306/allegheny-county-public-safety-assessment-court-28150/.

[117] *See* Email of November 17, 2016, from Paul Thomas, Court Administrator, Mesa Municipal Court, available at https://www.muckrock.com/foi/mesa-4736/mesa-municipal-court-public-safety-assessment-30126/; Letter of December 16, 2016 from Ann E. Donlan, Communications Director, Superior Court of California, County of San Francisco, available at https://www.muckrock.com/foi/san-francisco-city-and-county-3061/san-francisco-public-safety-assessment-court-30096/#file-113829; Email of February 6, 2017 from Marla Randall, Court Administrator, Navajo County Courts, available at https://www.muckrock.com/foi/navajo-county-9526/navajo-county-superior-court-public-safety-assessment-30129/; Letter of January 18, 2017 from Ronald G. Overholt, Court Administrator, Arizona Superior Court, Pima County, available at https://www.muckrock.com/foi/pima-county-183/pima-county-superior-court-public-safety-assessment-30130/#file-116730.

[118] *See, e.g.*, Memorandum of Understanding Between the Laura and John Arnold Foundation and the Superior Court of California, County of San Francisco, available at https://www.muckrock.com/foi/san-francisco-city-and-county-3061/san-francisco-public-safety-assessment-court-30096/#file-113830, at p.3 ("The Court agrees to refrain from disclosing, absent the entry of a court order by a court of competent jurisdiction, any information about the Tool, including information about the development, operation and presentation of the Tool, to any third parties without prior written approval from the Foundation.")

The one court system from which we received any documents about PSA other than a Memorandum of Understanding was the Seventh Judicial Circuit Court of Florida, on behalf of the Pretrial Services Program of Volusia County, one of the counties served by that circuit. This may be because Florida law requires private parties to expressly designate trade secrets or waive confidentiality – a feature of the law that was reflected in the MOU between the Arnold Foundation and the Seventh Judicial Circuit, which we also received.[119]

The documents produced by the Seventh Judicial Circuit provide some interesting additional information. For example, one document discloses the actual percentages of defendants, by risk score, who fail to appear or who commit new criminal activity or new violent crime. In what is apparently the original training data that Arnold used to create the algorithms, the percentages of defendants by risk score who were released and failed to appear are 1 (10%), 2 (15%), 3 (20%), 4 (31%), 5 (35%), 6 (40%).[120] Thus, the highest risk score was set to generate a risk of failure to appear four times that of the lowest risk score. Once the PSA algorithm started being used, however, the Arnold Foundation found that it generated a narrower band of results: 1 (12%), 2 (16%), 3 (18%), 4 (23%), 5 (27%), 6 (30%).[121] A score of "six" represented less risk of failure to appear than a score of "four" in the training data. Unfortunately, the only validation study results are three summary charts. Therefore, we have no way of knowing, for example, what percentage of the defendants in each risk category were detained rather than released before trial, and hence did not figure in the validation study.

Two documents produced by the Seventh Judicial Circuit also provide some information about another Arnold Foundation initiative that the Foundation itself has not broadly publicized. The Foundation recommends that courts use a "Decision Making Framework" which takes as input a defendant's PSA risk score and current pending charges, and generates as output specific recommendations as to pretrial treatment, from release without bail to detention. The Decision Making Framework is a second algorithm, generating specific recommendations for treatment (rather than risk scores). The Foundation states that Decision Making Frameworks are created for each jurisdiction by representatives of that jurisdiction in cooperation with a contractor that specializes in implementing the PSA program in particular court

---

[119] *See* Memorandum of Understanding Between the Laura and John Arnold Foundation and the Seventh Judicial Circuit of the State of Florida, available at https://www.muckrock.com/foi/volusia-county-10465/volusia-county-public-safety-assessment-court-28148/#file-112439 at 3.

[120] *See* Zach Dal Pra, LJAF Public Safety Assessment – PSA, available at https://www.muckrock.com/foi/volusia-county-10465/volusia-county-public-safety-assessment-court-28148/#file-112436, p. 31. This presentation, provided to us by the Seventh Judicial Circuit, contains only a brief summary of the study, with very little detail. Because it contains no citation to any published source, we assume that the study was conducted by the Foundation itself and has not been published.

[121] *Id.* p.46 (based on tracking PSA application in 100,000 cases in KY and in three unnamed cities outside of KY).

systems. However, the Seventh Judicial Circuit documents provide no information on how the Decision Making Framework for that court was created, or whether it has been subject to any testing.

Finally, we approached the Arnold Foundation directly and, through a series of emails and telephone conversations, asked specifically for technical reports, validation studies, and other documents the Foundation might have that would provide more detail about the creation and testing of the PSA algorithms. The Foundation responded with a short, three-page statement that consisted mostly of text that was already available on the Foundation's website.[122] We know from the Foundation's website, from the documents provided by the Seventh Judicial Circuit, and from the statement the Foundation produced for us, that the Foundation created the PSA algorithms by analyzing data in about 750,000 cases. We know nothing about how it analyzed that data, what alternatives it tried, or how those alternatives compared to the PSA algorithms it ultimately adopted.

We asked specifically why the Arnold Foundation had insisted on MOUs that prohibit courts from disclosing any information about PSA. The Foundation responded:

> Prior to releasing the algorithm, confidentiality agreements with early adopting jurisdictions kept PSA use limited while we developed local data infrastructure to measure results, waited for and studied post-implementation pretrial outcomes, and initiated additional research. These confidentiality agreements also helped to guard against the possibility of for-profit companies using elements of the PSA to develop substandard risk tools to be marketed to jurisdictions.

As far as we can tell, however, the confidentiality provisions are not limited to "early adopting jurisdictions," and the provisions all say that they require confidentiality in perpetuity.

2. Eckerd Rapid Safety Feedback – Child Welfare Assessments

Eckerd Rapid Safety Feedback (RSF) is a risk assessment process designed to identify child welfare cases with a high probability of serious child injury or death.[123] RSF was developed by Eckerd Kids (Eckerd), a nonprofit family and child services organization, and Mindshare Technology, a for-profit software company. Through a review of a large group of child welfare cases, including those in which children were injured or died, Eckerd identified the greatest risk factors contributing to child injury or death, namely, "a child under the age of three, a paramour in the home, substance abuse, and domestic

---

[122] That email exchange and attached document are available at http://www.robertbrauneis.net/algorithms/ArnoldFoundationEMail.pdf.

[123] *Summary and Replication Information*, ECKERD RAPID SAFETY FEEDBACK, available at http://static.eckerd.org/wp-content/uploads/Eckerd-Rapid-Safety-Feedback-Final.pdf

violence history, and a parent who had previously been placed in foster care."[124] Eckerd partnered with Mindshare Technology to create software that analyzes data in existing child welfare reporting systems and flags high-risk cases for intervention.[125]

We sent open records requests seeking information about use of the Eckerd RSF algorithm to five state child welfare agencies that Eckerd reported were using the RSF system: Alaska, Connecticut, Illinois, Maine, and Oklahoma. We received several documents from Alaska, Connecticut and Illinois. Oklahoma responded that it would need a payment of about $2500 to respond to our request, which would apparently include the cost of providing us the child welfare case data that it sent to Eckerd, with personally identifying information removed. Maine acknowledged our request but to date has not produced any documents.

The Alaska Department of Health and Social Services sent us a number of documents, including the Memorandum of Understanding between its Office of Child Services ("OCS") and Eckerd concerning Eckerd's provision of RSF assessments for child welfare cases to OCS. It is clear from the MOU that Eckerd retains control of the software that processes information about OCS child welfare cases and generates risk assessments. Child welfare case information is transmitted to Eckerd or Mindshare, and the risk assessments that are generated about those cases are made available on a website maintained by Eckerd, to which OCS personnel can gain access.[126]

The public agency, OCS, has no access to the algorithm that generates the risk assessments and none to the process by which the algorithm is generated and adjusted. Moreover, to the extent that OCS learns anything about the algorithm, it agrees not to disclose it. The Eckerd – Alaska OCS MOU provides that all "Eckerd IP," including the website maintained by Eckerd and its related software, reports generated by Eckerd, and all related inventions, processes, improvements and algorithms, are to be treated as "Confidential Information," which OCS agrees not to disclose.[127]

---

[124] *Id.*; *see also* Bryan Lindert, Eckerd Rapid Safety Feedback: Bringing Business Intelligence to Child Welfare, Policy & Practice, at 25, available at http://www.eckerd.org/wp-content/uploads/Eckerd.pdf. Eckerd also found that the most critical steps that can be taken to prevent child injury or death relate to quality safety planning, quality supervisory reviews, and the quantity and frequency of home visits. *Id.*

[125] *See Summary and Replication Information*, *supra* note x.

[126] *See* Memorandum of Understanding of February 20, 2015 between Eckerd Youth Alternatives, Inc. and the Alaska Department of Health and Social Services (available on the MuckRock website, https://www.muckrock.com/foi/alaska-235/alaska-department-of-health-and-social-services-eckerd-rapid-safety-feedback-software-documents-30807/#file-123161) p. 2 (providing that Eckerd will "[h]ost, maintain and support the Portal with a goal of providing the Agency with 24 hour technical support and access to the Portal and the reports it generates"); *id.* at 1 (defining "Portal" as "a website and related technology that is designed to read [electronic information about child welfare cases], perform automated analysis, and generate reports that can be used to implement and support Eckerd Rapid Safety Feedback.").

[127] *See id.* at 1, 4, 5.

The Connecticut Department of Children and Families provided a number of documents concerning Eckerd RSF, including a brochure, fact sheet, slide presentation, and flow chart, which confirm that the public agency provides information about child welfare cases to Eckerd, which then processes that information and generates risk assessments that the agency can view.[128]

The Illinois Department of Children and Family Services provided its contracts with Eckerd for Fiscal Year 2016 and Fiscal Year 2017. They show the amounts that Illinois estimates it will pay Eckerd for its services -- $107,000 in FY 2016 and $171,000 in FY 2017. The contracts also contain what appears to be standard state contracting provisions that are more favorable to disclosure and public ownership than the Alaska MOU. They recite that "[a]ny information not prohibited or exempt from disclosure under federal law, State law, or applicable FOIA exemption is public." They also provide that the state owns everything produced under the contracts, including all intellectual property rights and any products of the contracts. There is some language in the Illinois contracts suggesting that Eckerd is supposed to produce a new predictive algorithm based on analysis of Illinois data; one of the actions in "Phase 1: Development of the Model" is "The development of the predictive model that will be used to identify those incoming investigations with the highest probability of serious injury or death." It is unclear, however, whether this actually involves entirely new data analysis, or some fitting of an existing algorithm.

3. Allegheny Family Screening Tool – Child Welfare Assessments

Like Eckerd RSF, The Allegheny Family Screening Tool (AFST) was developed to facilitate the triaging of child welfare cases. AFST was developed by a consortium led by the Centre for Data Analytics at the Auckland University of Technology (the "Auckland Consortium"), in cooperation with the Allegheny County Department of Human Services. The Allegheny DHS published an RFP for projects to leverage Allegheny County's databases, and the Auckland Consortium submitted a successful proposal. While Eckerd RSF is apparently used on an ongoing basis to monitor cases within the child welfare system, AFST is applied at the time an initial call is made to report child maltreatment. It assists in determining whether the report warrants a formal investigation. Currently, Allegheny FST is used only in Allegheny County.

After we submitted an open records request to Allegheny County about the AFST, county officials contacted us, provided us with a report prepared by the Auckland Consortium about the development of the algorithm,[129] and indicated that they were

[128] For example, a chart entitled "CT Eckerd Rapid Safety Feedback Process Flow" (available on the MuckRock website, https://www.muckrock.com/foi/connecticut-53/connecticut-department-of-children-and-families-eckerd-rapid-safety-feedback-28152/#file-108367) allocates to Mindshare the step of "Mindshare Tool/Predictive Analysis Generates List for Review")

[129] *See* Rhema Vaithianathan, Emily Putnam-Hornstein, Nan Jiang, Parma Nand & Tim Maloney, Developing Predictive Models to Support Child Maltreatment Hotline Screening Decisions:

happy to speak with us about the algorithm and its development. The report is in many respects the most comprehensive we have seen on the development of an algorithm. It details many of the choices that were made in the development process, the reasoning behind those choices, and the data and methods that were used.

The developers ended up creating two algorithms, one for predicting the likelihood that an allegation, if not formally investigated, would be followed within two years by another allegation involving the same child, and another for predicting the likelihood that an allegation, if formally investigated, would result in the child being placed in foster care within two years. The training data for the algorithms was drawn from the County's integrated data management system, which was created in 2008; the developers decided that for each allegation of abuse in the dataset, they wanted data available for 18 months before that allegation, and two years after the allegation. The dataset included over 800 variables. The developers used nonlinear regression as their principal analytic method in large part because it produced as good results as other methods and had the advantage of being interpretable. In other words, it lent itself to transparency and accountability goals. The developers performed both internal validation studies – using a reserved portion of the training data – and external validation studies, using records of hospitalization and "critical events" (serious injury or death). The algorithm has not been in use long enough to conduct post-implementation studies.

The report discloses in an appendix 112 variables ultimately used – 71 for the model predicting foster home placement, and 59 for the model predicting repeat allegations – and the weights assigned to each of the variables are available upon request to the Allegheny DHS.[130] The output of the algorithms is presented as two risk scores – one for repeated allegations or "re-referral," and one for foster home placement – on a scale of 1 to 20, with each number representing a band of 5% of all children considered. Thus, a score of "10" would mean that the child's risk of re-referral or placement is in the 50-55% range of all children; a score of "15" would be in the 75-80% range. The developers also decided to create a threshold score that would presumptively result in a mandatory investigation, subject to the possibility of a supervisor waiving that outcome; the report does not disclose the threshold.

Allegheny County ultimately decided not to use the race of the child or custodians as a variable because it did not substantially improve predictive power and was otherwise problematic. The report discusses the dangers of false negatives and false positives at some length, but does not say whether they were ultimately weighted equally or unequally.

Allegheny County Methodology and Implementation (2017), http://www.alleghenycountyanalytics.us/wp-content/uploads/2017/04/Developing-Predictive-Risk-Models-package-with-cover-1-to-post-1.pdf.

[130] We made such a request and were provided with the weights.

Although the Auckland Consortium has retained copyright in the code used to implement the algorithm, the contract with Allegheny County grants it the power to license other jurisdictions to use that code without further payment, and county officials have indicated that they are interested in doing so. Thus, although this project is not fully an open source project, it comes closer than any of the other five algorithms we studied.

4. PredPol – Predictive Policing

PredPol is software that predicts where and when crimes of various types are likely to occur, and thus assists police forces in plotting their patrols to deter those crimes. It was originally developed by mathematicians and behavioral scientists from UCLA and Santa Clara University in collaboration with crime analysts and officers from the Los Angeles and Santa Cruz Police Departments, [131] but is now managed by a for-profit company, PredPol Inc. The creators of PredPol determined that the three most important types of information or "data points" for predicting crime are crime type, crime location, and crime date and time.[132] PredPol feeds data about past patterns of criminal activity into an algorithm that predicts where and when new crimes will be committed.[133] According to one source, PredPol "is well known for keeping its algorithm 'a closely guarded secret.'"[134]

We sent requests for records concerning PredPol to eleven police departments, including the police departments of Oxford, Alabama; Little Rock, Arkansas; Los Angeles, Modesto, Orange County and Santa Cruz, California; Cocoa, Florida; Atlanta, Georgia; Hagerstown, Maryland; Reading, Pennsylvania; and Tacoma, Washington. Eight of those eleven police departments either did not respond, or acknowledged our request but did not produce documents, or asked for more time to respond and have not yet responded, or responded that they did not have any relevant documents. The three departments that did provide documents were those of Tacoma, Cocoa, and Santa Cruz.

The City of Tacoma, Washington was among the most forthcoming of any of the jurisdictions to which we sent records requests about any algorithm. It supplied 200 email threads of correspondence between Tacoma Police Department and PredPol personnel concerning a wide variety of issues in implementing PredPol. It also produced 10 presentations on how PredPol and predictive policing work. These documents would be quite helpful to someone interested in what PredPol reports look like, what data the PredPol algorithm uses as input, and so on. None of the documents,

---

[131] *PredPol is Predictive Policing*, PREDPOL, *available at* http://www.predpol.com/about/ (last visited March 19, 2017).

[132] *Id.*

[133] *Id.*

[134] Joh [2017], supra note xx at 19 (quoting Ali Winston, Arizona Bill would fund predictive policing technology, REVEAL, Mar. 25, 2015, at https://www.revealnews.org/article/arizona-bill-would-fund-predictivepolicing-technology/).

however, reveals the algorithm that PredPol is using to generate predictions from past crime data, nor the process that PredPol used to create that algorithm.

The City of Cocoa sent us a number of documents that all related to the purchase of services from PredPol. Perhaps most telling was the background document provided to City Council members when the purchase of PredPol services was on the Council agenda. That document does not provide any detail about PredPol, but states that "[t]he City Attorney has advised that information revealing surveillance techniques, procedures or personnel is exempt from public inspection pursuant to s. 119.071(2)(d), Florida Statutes." It is likely that the City relied on this advice in declining to provide any documents about PredPol itself, although it is very likely that the city could conceal surveillance techniques while still being more transparent about the algorithm's values and implementation.

The City of Santa Cruz, California sent several screenshots of PredPol software. One screen requests the user to input data about the place (in latitude and longitude), time, and type (vehicle or residential) of recent crimes, and states that predictions about the location of crimes over the following 24-hour period will appear on a map. The other screen is a map of the City, with colored areas representing where crimes are likely to occur. Those screenshots provide information about the type of data input and the format of the output, but little else. The PredPol version that Santa Cruz is using appears to be less sophisticated than that used by Tacoma.

5. HunchLab – Predictive Policing

Like PredPol, HunchLab is software that predicts where and when crime will occur, with a cartographic output indicating areas at higher risk for certain types of crimes over certain time periods. HunchLab is developed and maintained by Azavea, Inc., a for-profit certified B corporation.[135] HunchLab uses a wide range of inputs to predict risks of crime, and allows individual police departments to prioritize for selected crimes.[136]

We sent open records requests concerning HunchLab to four police departments, including those of Miami, Florida; St. Louis County, Missouri; Lincoln, Nebraska; and Philadelphia, Pennsylvania.

---

[135] B corporations are for-profit corporations certified by B Lab, a nonprofit certification organization, to meet certain standards of social and environmental performance, accountability, and transparency. *See* What are B Corps?, https://www.bcorporation.net/what-are-b-corps. [https://www.themarshallproject.org/2016/02/03/policing-the-future#.PBX0B3Cac - story on HunchLab deployed in St. Louis]

[136] "HunchLab determines what data is most useful for prediction of each crime. In some cases, geography — the locations of prior crimes or particular landmarks — is the most important factor. In others, time — day of week, month of year — takes precedence." https://www.themarshallproject.org/2016/02/03/policing-the-future#.vVL53xF4m (reporting on St. Louis' deployment of HunchLab system).

Miami never responded to our initial request, or to a follow-up request.[137] St. Louis County asked for a payment of $400 before it produced any documents, and reiterated that it would not act without a $400 payment when we asked whether we could narrow our request to reduce the fee.[138] The City of Philadelphia produced a purchase order for the HunchLab service,[139] but otherwise denied our request on the grounds that we did not request specific documents.[140] The City of Lincoln, Nebraska provided several documents, including a manual introducing HunchLab to staff, and a blog post by the City's Public Safety Director on HunchLab. Perhaps most helpfully, Lincoln provided us with a sample set of input data for HunchLab, which it identified as comprising a 30-day rolling window of police incident reports.[141] Over that time period, Lincoln recorded 3057 police incident reports; each of those reports contains details about the street address, latitude and longitude of the reported crime; the type of crime; and the date and time of the report and of the crime.[142]

Jeremy Heffner, HunchLab Project Manager and Senior Data Scientist at Azavea, approached us after learning of our open records request to Lincoln, Nebraska. We had an email exchange and phone conversation with him, and he ultimately created and sent us a draft document titled "A Citizen's Guide to HunchLab," which provides information about the HunchLab algorithms and their creation and validation. It seems from that document that the HunchLab algorithms are less interpretable than many others. They are built using a "random forest" technique in which successive decision trees are tried and tested; the developers incorporate data, not just about reported crimes and the place and time they occurred, but data about location of known offenders, location of known and likely targets of crime, weather, daily, weekly, and seasonal cycles, socioeconomic indicators, and so on.[143] A police officer cannot know how the algorithm's decisionmaking relates to his or her own knowledge and judgment. The HunchLab algorithm is also the most dynamic of any of the algorithms we studied. For each client, HunchLab does what it calls a "new modeling run" every

---

[137] *See* https://www.muckrock.com/foi/miami-103/miami-pd-hunchlab-documents-30109/ (displaying correspondence).

[138] *See* https://www.muckrock.com/foi/st-louis-county-8838/st-louis-county-pd-hunchlab-documents-30113/ (displaying the exchange of correspondence).

[139] *See* https://www.muckrock.com/foi/philadelphia-211/philadelphia-pd-hunchlab-documents-30111/#file-119920.

[140] *See* Letter of February 26, 2017 from Robert Kieffer, Assistant City Solicitor, available at https://www.muckrock.com/foi/philadelphia-211/philadelphia-pd-hunchlab-documents-30111/#file-119922.

[141] *See* Letter of November 30, 2016 of Tonya Peters, available at https://www.muckrock.com/foi/lincoln-4033/lincoln-police-department-hunchlab-documents-30110/#file-110327.

[142] See the data files available for download at https://d3gn0r3afghep.cloudfront.net/foia_files/2016/11/30/Archive.zip.

[143] *See* A Citizen's Guide to HunchLab 12 (Draft July 11, 2017).

few weeks to re-calibrate the model, and each of those modeling runs creates a new predictive algorithm.[144]

HunchLab also discusses openly the issue of potential bias in inputs, and its judgments on that issue. One type of bias is "reporting bias" – some communities may report larger percentages of crimes than others. HunchLab takes the position that it is appropriate to incorporate much of that bias into police activity. It states: "We believe that police activity should reflect what the community is reporting as problems. . . . If reporting biases are due to distrust of the police, then we believe that letting the bias exist within the data is appropriate."[145] It notes that this may not be the case if failure to report is due to fear or shame, but it is not clear how that can be remedied. HunchLab also comments on "enforcement bias" – the possibility that police end up making more arrests and engaging in more enforcement activity in some communities than in others, even if the level of crime is the same. It states a belief that that bias is less present in major crimes such as homicides, robberies, or assaults; for other drug-related and nuisance crimes, it states that it tries to use data that reflects the community's call for services – complaints – rather than data that reflects police enforcement activity.[146]

The HunchLab program has three other interesting features. First, the algorithm allows each community to set weights for the relative seriousness of each type of crime – how much more important is it to stop a murder than a burglary? It also allows tailored weights for patrol efficacy – indoor crimes are less likely to be deterred by increased police presence.[147] Second, HunchLab recommends that the algorithm incorporate randomness to assure that police are not assigned to the same routes every day, to combat monotony on the job, and to reduce the negative side effects of constant police presence in an area.[148] Third, HunchLab has now extended its reach into patrol tactics, recommending certain kinds of police activity in patrol areas, such as car patrol, foot patrol, car stops, etc., and over time monitoring the effectiveness of the tactics used.[149]

6. New York City and New York State Value Added Models – Teacher Evaluation

New York City and the State of New York are among the jurisdictions that have adopted a Value Added Model ("VAM") method for evaluating teachers.[150] In general,

---

[144] *Id.* at 19.
[145] *Id.* at 2.
[146] *Id.* at 26.
[147] *Id.* at 9-10.
[148] *Id.* at 10-11.
[149] *Id.* at 11-14.
[150] The New York Supreme Court held that the New York City growth measurements were arbitrary and capricious as to the complaining teacher. Matter of Lederman v King 2016 NY Slip Op 26416 (May 10, 2016). *See generally*, https://www.washingtonpost.com/news/answer-sheet/wp/2016/05/10/judge-calls-evaluation-of-n-y-teacher-arbitrary-and-capricious-in-case-against-new-u-s-secretary-of-education/.

Value Added Model algorithms compare test scores of students at the beginning and end of a given year in order to measure the progress of those students. Those results are then adjusted to try to account for factors other than teacher effectiveness, such as socioeconomic status, that might be responsible for the students' progress or lack thereof. The adjusted results for the students that are taught by a particular teacher are then used to produce an evaluation of that teacher's effectiveness.

We filed open records requests with both the City of New York and New York State for documents relating to their VAM programs.[151] To date, the City of New York has sent us five letters notifying us that it needs more time to produce records, but it has not sent us any records.[152] The New York State Education Department produced a number of documents, including the original contract with its vendor, the American Institute of Research, to implement a VAM program for New York; two renewals of that contract; five published articles by various authors generally evaluating the validity of Value Added Models, none of which focuses on the New York VAM implementation; and sample outputs of the VAM algorithm – outputs for 50 students and 50 teachers, with their names and other identification removed.

The sample outputs do provide some information about the format of what the VAM algorithm produces, and they provide a glimpse of how the algorithm works, because they actually contain some of the inputs – for example, student test scores – as well as the outputs. However, 50 sample outputs is much too small a number to begin to reverse engineer the algorithm, and the contract between the Education Department and the American Institute of Research provides that "methodologies or measures that are the property of the contractor at the time the contract is executed" are "proprietary information" that the Education Department is allowed to use "solely for [its] educational purposes."[153] Thus, the algorithm or algorithms are not publicly available, and the process by which they were constructed has not been disclosed.

Conclusion

Our efforts to learn about predictive algorithms through open records requests were in many respects frustrating. Many governments did not respond, and many of those that did claimed to be either generally exempt from open records acts (as, for example, courts) or beneficiaries of specific exemptions, such as those for trade secrecy. While a number of jurisdictions provided their contracts with vendors, thus enabling us to learn something about contract terms, we got very little about the development of the algorithms, probably because the governments were never in possession of records that

[151] Ours was not the first attempt. Cathy O'Neil also tried and failed to obtain New York City's VAM records. https://mathbabe.org/2014/03/07/an-attempt-to-foil-request-the-source-code-of-the-value-added-model/

[152] *See* https://www.muckrock.com/foi/new-york-city-17/new-york-state-or-new-york-city-value-added-measures-for-teachers-29739/ (displaying correspondence).

[153] *See* Contract No. C010834, between the People of the State of New York and American Institutes of Research, dated September 19, 2011, Appendix D, available at [post contract].

would include that information. Allegheny County, which contracted for the development of a predictive algorithm from scratch by a consortium of university researchers, was the biggest exception, because it commissioned and possessed reports that detailed the development of its algorithm and disclosed the algorithm itself.

III. Obstacles to Transparency

Having detailed our efforts to obtain useful information through open records requests about state and local government deployment of predictive algorithms, we now turn to the obstacles we encountered. Principal among these were a failure to generate or deliver to government important records, and claims of trade secrecy. We also discuss the law enforcement and deliberative process exemptions in open records laws which are likely to be overused because governments fear algorithmic transparency.

A. Lack of Documentation

Governments cannot disclose more information than they have. Most open records laws entitle requesters only to obtain "records" or "information" already in existence. Agencies are generally not required to generate new records when faced with an open records request.[154] Our research suggested that governments simply did not have many records concerning the creation and implementation of algorithms because those records were never generated, or were generated by contractors and never provided to the governmental clients. These include records about model design choices, data selection, factor weighting, and validation designs. At an even more basic level, most governments did not have record of what problems the models were supposed to address, and what the metrics of success were.

Many of the most important decisions in a big data application are made at the "wholesale" level of the design of a model, not at the "retail" level of application to a particular case. In the analog world, wholesale policy decisions that are not legislated are likely to be made through administrative rulemaking. There is the announcement of a proposed policy, opportunities to comment on that proposal, and eventually disclosure of the final policy, the reasons why it was adopted, and an explanation of how it will be implemented. These norms and laws do not apply to the creation of algorithmic policy. Big data prediction models are often built and used without key policy decisions ever having been articulated or justified. In the best cases, there will be public requests for proposals for private vendors to supply predictive algorithms to government.[155] More typically, there will simply be a form agreement with a private

---

[154] *See, e.g.*, Calif. Public Records Act: Gov't Code §6250-6268; Guidance at http://www.thefirstamendment.org/publicrecordsact.pdf ("The PRA covers only records that already exist, and an agency cannot be required to create a record, list, or compilation.").

[155] *See, e.g.*, Allegheny County Department of Human Services, Request for Proposal to Design and Implement Decision Support Tools and Predictive Analytics in Human Services (2014), http://www.county.allegheny.pa.us/Human-Services/Resources/Doing-

vendor that does not articulate the political choices that have been embedded in the algorithm.

B. Aggressive Trade Secret and Confidentiality Claims

Even where governments have key explanatory records, they may refuse to disclose them in deference to the claims of private vendors that this information is confidential. The owners of proprietary algorithms will often require nondisclosure agreements from their public agency customers[156] and assert trade secret protection over the algorithm and associated development and deployment processes.[157] Governments will then use these claims to exempt vendor material from disclosure, often in ways that violate the open records laws' relatively narrow trade secret exemptions.

As discussed above, we encountered jurisdictions that cited trade secrets and confidentiality as reasons they could not reveal more about their predictive models. This was true, for example, of the Mesa Municipal Court and the Pima and Navajo County Court systems in Arizona, and the San Francisco Superior Court system in California, who were using the Arnold Foundation's PSA-Court.[158] It was also true of Alaska, using the Eckerd Rapid Safety Feedback risk assessment for children.[159] The open records laws of Arizona,[160] California,[161] and Alaska[162] all exempt trade secrets and confidential information, and none entitles public access to records the government does not have. It would require considerable more probing and perhaps litigation to determine if the government agencies acted lawfully. But what we can say is that the agencies have agency. They could have made more records disclosable simply by reducing the scope of confidentiality and ensuring government possession of records necessary to explain the algorithms.[163]

Business/Solicitations/2014/Decision-Support-Tools-and-Predictive-Analytics-in-Human-Services-RFP.aspx.

[156] *See* Elizabeth E. Joh, The Undue Influence of Surveillance Technology Companies on Policing (February 27, 2017). __ N.Y.U. L. Rev. Online __ at 7 (forthcoming 2017) (discussing police department nondisclosure agreements with the Harris Corporation for use of Stingray police surveillance technology), https://ssrn.com/abstract=2924620.

[157] *See generally* Kitchin, Thinking Critically about and Researching Algorithms, *supra* note 5, at 20. *See e.g.*, Houston Federation of Teachers, *supra* note57, at 12 (teacher evaluation "scores are generated by complex algorithms, employing 'sophisticated software and many layers of calculations.'" The vendor "treats these algorithms and software as trade secrets, refusing to divulge them either to [the District] or the teachers themselves.").

[158] *See supra* n. 136.

[159] *See supra* n. 147.

[160] Arizona Public Records Law, A.R.S. § 39-121.

[161] California Public Records Act, Cal. Govt. Code §§ 6250 to 6276.48.

[162] Alaska Public Records Disclosures, AS §§ 40.25.100 to 350.

[163] As noted above, Florida's Seventh Judicial District took a step in the right direction, *see supra* n. 119, but it could have done much more.

Overbroad assertions of confidentiality in response to open records requests are common in the field. For example, in researching how California police departments use the Shotspotter technology to respond to gunshots fired in their jurisdictions, Forbes reporter Matt Drange submitted more than a dozen state freedom of information act requests for Shotspotter-generated reports of gunfire.[164] Despite the fact that the requests did not seek the underlying sensor technology, the jurisdictions initially reported that they could not disclose the data as a result of confidentiality agreements with Shotspotter.[165] Risk-averse municipalities thought they could not share information on shots detected in their jurisdictions even though the data was not a trade secret or confidential.

The roadblocks to algorithmic transparency are especially problematic in the criminal justice context where individual liberty is at stake. Journalists were unsuccessful in obtaining information about NorthPointe's COMPAS ("Correctional Offender Management Profiling for Alternative Sanctions") sentencing algorithm through open records requests because of alleged trade secret protection.[166] In litigation related to the algorithm, a Wisconsin appellate court upheld use of COMPAS against a defendant's due process claim, but acknowledged the transparency problem and required that sentencing reports inform judges that "the proprietary nature of [the algorithm] has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are determined."[167]

Government assertions of trade secrecy protection on behalf of their vendors may sometimes be justified. Government agents are subject to ordinary liability for disclosing trade secrets and/or for violating nondisclosure agreements, unless protected by some form of immunity.[168] Most states have adopted the Uniform Trade

---

[164] Matt Drange, "We're Spending Millions On This High-Tech System Designed To Reduce Gun Violence. Is It Making A Difference?", *Forbes* (Nov. 17, 2016), http://www.forbes.com/sites/mattdrange/2016/11/17/shotspotter-struggles-to-prove-impact-as-silicon-valley-answer-to-gun-violence/#27e6920c9dbf.

[165] The company had sent out a nationwide memo to customers in July 2015, urging cities to issue blanket denials to records requests or disclose heavily redacted information, "in a form that would not harm SST's business and allow the customer to respond from a public goodwill point of view." https://www.documentcloud.org/documents/3221020-ShotSpotter-nationwide-memo-July-2015.html. *See also* Jason Tashea, Should the Public Have Access to Data Police Acquire Through Private Companies?, ABA J., DEC. 1, 2016, at http://www.abajournal.com/magazine/article/public_access_police_data_private_company.

[166] Angwin *et al*, *supra* note 6; *see also* Diakopolous, *supra* note 8.

[167] *State v. Loomis*, 371 Wis.2d 235, 881 N.W.2d 749 (2016). It also required disclosure that no validation studies have been completed and tools must be constantly monitored and re-normed.

[168] *See, e.g.*, RCW 42.56.060 ("No public agency, public official, public employee, or custodian shall be liable, nor shall a cause of action exist, for any loss or damage based upon the release of a public record if the public agency, public official, public employee, or custodian acted in good faith in attempting to comply with the provisions of this chapter."). *Accord Levine v. City of Bothell*, 2015 WL 2567095 at *3 (W.D. Wash. 2012) (recognizing that "a public agency and its employees are immune from liability upon the release of public records if they acted in good

Secret Act,[169] which protects against the "misappropriation" of a trade secret, defined as "disclosure or use of a trade secret of another without express or implied consent by a person who…at the time of disclosure or use, knew or had reason to know that his knowledge of the trade secret was…acquired under circumstances giving rise to a duty to maintain its secrecy or limit its use".[170] Governments are persons.[171] When a private vendor asserts trade secret protection, and further demands that government officials sign nondisclosure agreements, this creates a pull towards secrecy. It is a pull strengthened by the government agency's own interests in secrecy, for reasons discussed below. The countervailing pull towards transparency comes only from open records acts and other transparency policies.

By exempting trade secrets from disclosure, open records acts are not the force for transparency that they might otherwise be.[172] In most states, the exemption is express.[173] Exemption 4 of FOIA is has many close parallels in state open records act exemptions. It excludes from disclosure "trade secrets and commercial or financial information obtained from a person [that is] privileged or confidential."[174] The

---

faith by attempting to comply with" the Washington open records law"). *Cf.* Peter S. Menell, *Tailoring a Public Policy Exception to Trade Secret Protection*, 105 CAL. L. REV. 1, 30-31 (2017) (discussing privileges for private parties to disclose trade secrets in the public interest).

[169] Unif. Trade Secrets Act, 14 U.L.A. 71 table of jurisdictions adopting act (1985) [UTSA].

[170] UTSA §(1)(2)(ii).

[171] *Id.* at §(1)(3). In addition, federal law specifically forbids disclosure of trade secrets. 18 U.S.C. §1905 (imposing criminal liability on any US government employee who in course of official duties, "publishes, divulges, discloses, or makes known in any manner or to any extent not authorized by law any information… which information concerns or relates to the trade secrets, processes, operations, style of work, or apparatus, or to the identity, confidential statistical data, amount or source of any income, profits, losses, or expenditures of any person").

[172] *See, e.g.*, Conn. Gen. Stat § 1-210(b)(5)(exempting from disclosure a "trade secret," defined as (A) "information, including formulas, patterns, compilations, programs, devices, methods, techniques, processes, drawings, cost data, or customer lists that (i) derive independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable by proper means by, other persons who can obtain economic value from their disclosure or use, and (ii) are the subject of efforts that are reasonable under the circumstances to maintain secrecy; and (B) Commercial or financial information given in confidence, not required by statute."); Del. Code Ann. 29 § 10002(g)(2) ("Trade secrets and commercial or financial information obtained from a person which is of a privileged or confidential nature" are deemed not to be public); Pennsylvania Right to Know Act, § 101 to 3104, § 708(b)(11) ("A record that constitutes or reveals a trade secret or confidential proprietary information" is exempt from disclosure.). *See generally*, The Reporters Committee for Freedom of the Press, Open Government Guide, http://www.rcfp.org/ogg/index.php (compendium of all 50 state open records acts). *See generally*, Linda B. Samuels, *Protecting Confidential Business Information Supplied to State Governments: Exempting Trade Secrets from State Open Records Laws*, 27 Am. Bus. L.J. 467, 468-69 (1989).

[173] *See supra* note xx. In other states, courts recognize trade secrets under more general exemptions. *See, e.g.*, Phoenix Newspapers, Inc. v. Keegan, 201 Ariz. 344, 351 (Ariz. Ct. App. 2001) (recognizing that trade secrets are "protected by the confidentiality exception to disclosure" in Arizona's open records law).

[174] 5 U.S.C. § 552(b)(4) (2016).

exemption thus covers two broad categories: (1) trade secrets and (2) information that is (a) commercial or financial and (b) obtained from a person, and (c) privileged or confidential.

How much trade secret exemptions actually contract the transparency mandate of open records laws depends on agency and judicial interpretations. When challenged, overly generous agency protections have been struck down. The D.C. Circuit - the leading source of FOIA case law - has interpreted the term "trade secret" to have a more limited meaning than it does under the Uniform Trade Secrets Act[175] (and the federal Defend Trade Secrets Act of 2016).[176] The government may only withhold records under Exemption 4 for "a secret, commercially valuable plan, formula, process, or device that is used [in connection with] trade commodities and that can be said to be the end product of either innovation or substantial effort."[177] There must be "a direct relationship between the information at issue and the productive process," rather than merely "collateral business confidentiality."[178] In other words, the information concealed must be central to the commercial product, and not merely an ancillary byproduct. Given this limitation, not all algorithmic processes that a vendor might consider to be a trade secret in the commercial sphere will count as a trade secret for open records exemption purposes.

The second prong of Exemption 4 permits secrecy for some kinds of commercial information.[179] This part of the exemption is also limited. The information has to be "privileged or confidential."[180] The D.C. Circuit has held that a mere promise of confidentiality to the source of the information is insufficient. Rather, the government

---

[175] UTSA §1.4 ("information, including a formula, pattern, compilation, program, device, method, technique, or process, that: (i) derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable by proper means by, other persons who can obtain economic value from its disclosure or use, and (ii) is the subject of efforts that are reasonable under the circumstances to maintain its secrecy.")

[176] 18 U.S.C. § 1839(3) ("all forms and types of financial, business, scientific, technical, economic, or engineering information, including patterns, plans, compilations, program devices, formulas, designs, prototypes, methods, techniques, processes, procedures, programs, or codes, whether tangible or intangible, and whether or how stored, compiled, or memorialized physically, electronically, graphically, photographically, or in writing.")

[177] Pub. Citizen Health Research Group v. U.S. Food and Drug Admin., 704 F.2d 1280, 1288 (D.C. Cir. 1983); Id. at 1289 ("The Restatement approach, with its emphasis on culpability and misappropriation, is ill-equipped to strike an appropriate balance between the competing interests of regulated industries and the general public."). *See also* Anderson v. Dep't of Health & Human Servs., 907 F.2d 936, 944 (10th Cir. 1990) (adopting the same definition).

[178] Pub. Citizen Health Research Grp., 704 F.2d at 1288, 1287.

[179] DOJ FOIA GUIDE, supra note xx, at 266–67 (citing Pub. Citizen Health Research Grp., 704 F.2d at 1290)(records are commercial if the submitter "has a 'commercial interest' in them."). *See also*, Critical Mass Energy Project v. Nuclear Regulatory Comm'n, 975 F.2d 871, 872–74 (D.C. Cir. 1992).

[180] HARRY A. HAMMITT ET AL., LITIGATION UNDER THE FEDERAL OPEN GOVERNMENT LAW 2010 119 (25th ed. 2010) (The term "confidential" is "the key term in Exemption 4 caselaw.").

must prove that disclosure would likely (1) "impair the government's ability to obtain necessary information in the future" or (2) "cause substantial harm to the competitive position" of the information source.[181] Given the burden that the government bears, some states require that state agencies notify private entities of requests for trade secret or confidential information and obtain the private party's defense of its designation.[182] Ultimately, it is the state that makes the call.

In interpreting trade secrets exemptions, government officials should be mindful of the purpose of the carve-out. The purpose of FOIA Exemption 4 is to preserve the government's ability to collect information from regulated entities, [183] or in an alternative formulation, to "encourage individuals to provide certain kinds of confidential information to the Government . . ."[184] Similarly, state open records laws protect trade secrets and confidential information *in order to* advance public goals.[185] Because open records laws impose a presumption of openness, and because states have followed FOIA courts in construing trade secrets and confidential material narrowly, the trade secret exemption to open records is narrower than private vendors might like. This is especially true when government is acting as a customer and not as a regulator. Currently pending litigation in New York poses the question of how far trade secret claims should be honored when government acts in its enterprise capacity. Citing trade secret protection, [186] New York City refused the Brennan Center for Justice's freedom

---

[181] Nat'l Parks and Conservation Ass'n v. Morton, 498 F.2d 765 (D.C. Cir. 1974). [D.C. Cir. later revisited and retained en banc - cite]

[182] *See, e.g.*, PA Right to Know Law Exemption 708 (state agency must notify a company of a request to disclose trade secret or confidential information within five business days. The company then has five business days to provide the state agency with the company's position concerning disclosure of its information. Within 10 days of notifying the company, the state agency must decide to release or withhold the information); [NY law]; [TX law]; [MS law].

[183] According to the U.S. Justice Department, Exemption 4 "affords protection to those submitters who are required to furnish commercial or financial information to the government by safeguarding them from the competitive disadvantages that could result from disclosure." U.S. DEP'T OF JUSTICE, GUIDE TO THE FREEDOM OF INFORMATION ACT 263, https://www.justice.gov/oip/foia-guide-2004-edition-exemption-4. *See also* Attorney General's Memorandum for Heads of All Federal Departments and Agencies Regarding the Freedom of Information Act (Oct. 12, 2001), reprinted in *FOIA Post* (posted 10/15/01) (recognizing fundamental societal value of "protecting sensitive business information").

[184] Soucie v. David, 448 F.2d 1067, 1078 (D.C. Cir. 1971).

[185] *See e.g.*, Verizon N.Y., Inc. v. N.Y. State Pub. Serv. Comm'n, 137 A.D.3d 66, 71 (N.Y. App. Div. 3d Dep't 2016) ("policy behind Public Officers Law § 87(2)(d) is simply to protect businesses from the deleterious consequences of disclosing confidential commercial information, so as to further the State's economic development efforts and attract business to New York.").

[186] N.Y. Pub. Off. Law § 87(2)(e) (iv) (exempting from disclosure records that "are trade secrets or are submitted to an agency by a commercial enterprise or derived from information obtained from a commercial enterprise and which if disclosed would cause substantial injury to the competitive position of the subject enterprise.")

of information law requests for records related to the sentencing algorithm known as Palantir Gotham.[187]

All of the requests we made were submitted to jurisdictions acting in their enterprise capacity. We can be confident that the assertions of trade secret over all materials connected with the algorithms were overbroad. Even assuming that the source code and certain details of the model would qualify as trade secrets or confidential information, we sought training materials, existing and planned validation studies, and other documentation concerning the objectives and design choices reflected in the algorithm. It is hard to imagine that most if any of this material would qualify for the exemption.

It is almost certainly true that protecting algorithms as trade secrets sometimes incentivizes companies to create predictive models for public applications.[188] At the same time, the information allegedly protected by trade secret law may lie at the heart of essential public functions, and constitute political judgments long open to scrutiny. As David Levine writes, "[t]he conflict between trade secrecy and a transparent and accountable democratic government is ultimately a clash of governing theory and values."[189] It is a conflict that can be mitigated by courts and legislatures limiting the scope of the trade secret exemption to open records laws and by government agencies insisting on transparency when they contract for algorithms.

C. Other Governmental Concerns and Open Records Act Exemptions

Even if government agencies generated or acquired sufficient records and assured that those records were not subject to claims of trade secrecy, they might have other reasons for resisting algorithmic transparency: gaming or circumvention; loss of candor in deliberation; and undue public controversy.

Government officials may worry that publicly-disclosed algorithms will be gamed or circumvented, making predictions less reliable and thwarting their purpose.[190] If a

---

[187] https://www.brennancenter.org/sites/default/files/8%20-%20Memorandum%20of%20Law%20in%20Support%20of%20Verified%20Petition.pdf.

[188] Kroll et al., *supra* note 25 at 15-17. *See also* David S. Levine, *Secrecy and Unaccountability: Trade Secrets in Our Public Infrastructure*, 59 Fla. L. Rev. 135, 180-181 (2007) (providing an example in the voting machine context of how state laws compelling source code disclosure can deter companies from contracting with the state for public services).

[189] *See* Levine, *supra note 188*, at 157; *see also* Mark Fenster, *The Opacity of Transparency*, 91 Iowa L. Rev. 885, 919 (2006) (observing a "fundamental conflict between laws intended to cover government agencies and the increasing reliance by those agencies on private firms" and noting that state courts and legislatures have "failed to develop a consensus or clarity for their open government laws" to address this conflict).

[190] In machine learning literature, the gaming problem is known more generally as "adversarial learning" – the problem of developing models when it is anticipated from the beginning that adversaries will try to defeat them. *See, e.g.*, Daniel Lowd & Christopher Meek, "Adversarial learning," in Proceedings of the Eleventh ACM SIGKDD International Conference on

criminal defendant knows that statements she makes will result in a higher recidivism risk score, she may lie.[191] If a terrorist knows how names are placed in the Terrorist Screening Database and matched to names on visa applications, he may try to avoid such placement and matching.[192]

These concerns are understandable, but do not excuse nonresponsiveness to open records requests. Open records acts do address potential gaming in the context of law enforcement investigations and investigative techniques.[193] Exemption 7(E) of FOIA asks explicitly whether that disclosure of investigative techniques would "risk circumvention of the law."[194] However, this exemption does not cover predictive policing programs like PredPol and HunchLab, which do not concern "investigations" in the core sense of gathering evidence of already-committed crimes. Some courts have been willing to stretch "investigation" to cover preventative measures.[195] One of our open records requests revealed that one jurisdiction exempted itself from providing

---

Knowledge Discovery and Data Mining (KDD) 641 (A. Press, ed., 2005); Pavel Liskov & Richard Lippmann, *Machine Learning in Adversarial Environments*, 81 Machine Learning 115 (2010).

[191] For example, COMPAS, a tool for assessing the likelihood of recidivism by criminal defendants, bases its predictions in part on a defendant's agreement or disagreement with statements such as "A hungry person has the right to steal" and "You can talk your way out of a problem." *See* Brittney Via, Amy Dezember & Faye S. Taxman, *Exploring How to Measure Criminogenic Needs: Five Instruments and No Real Answers*, in Handbook on Risk and Need Assessment: Theory and Practice (Faye S. Taxman, ed. 2017).

[192] *See* Jerome P. Bjelopera, Bart Elias & Aaron Siskin, The Terrorist Screening Database and Preventing Terrorist Travel (CRS Report R44678) 12 (Nov. 7, 2016) (documenting the use of name-searching algorithms in screening visa applicants). For an exploration of the fuzzy line between enforcement and prevention, see Coglianese & Lehr, *supra* note 52, at 1210 (an algorithmic rulemaking process might model compliance choices of regulated entities, in which case it would be similar to post-hoc enforcement algorithms and might legitimately be exempted from disclosure).

[193] *See, e.g.*, 5 U.S.C. § 552(b)(7)(E) (exempting "records or information compiled for law enforcement purposes" whose disclosure "could reasonably be expected to risk circumvention of the law."); 5 Ill. Comp. Stat. 140/7(d)(v) (exempting law enforcement records that "disclose unique or specialized investigative techniques other than those generally used . . ."); Mich. Comp. Laws §15.243(1)(b)(v) (exempting records that would "[d]isclose law enforcement investigative techniques or procedures").

[194] 5 U.S.C. § 552(b)(7)(E).

[195] *See* Coastal Delivery Corp. v. United States Customs Service, 272 F.Supp.2d 958 (C.D. Cal. 2003) (holding that the Customs Service could withhold records of the number of examinations of merchandise arriving into various seaports under Exemption 7(E), because they could aid the illegal importation of goods by informing importers of where and when examinations were less likely to occur); U.S. News & World Report v. Dept. of the Treasury, 1986 U.S. Dist LEXIS 27634 (D.D.C.) (holding that details of construction of the President's limousines could be withheld under Exemption 7(E), and adopting a broad reading of "investigative" that encompassed preventing potential harm to the President); *but see* Living Rivers, Inc. v. United States Bureau of Reclamation, 272 F.Supp.2d 1313, 1320-1322 (D. Utah 2003) (holding that maps of areas below dams that would be inundated if the dams were breached could not be withheld under Exemption 7(E), because the maps did not disclose investigative practices).

data related to police surveillance techniques, arguably a cousin of prevention.[196] But crime prevention methods and their relatives are at or beyond the periphery of the exemption. Risk assessment of criminal defendants for recidivism and failure to appear seems even less tied to "investigation." Moreover, there is no exemption from open records laws for other non-criminal justice gaming concerns. Child welfare programs like the Eckerd Rapid Safety Feedback and Allegheny Family Screening Tool efforts are not primarily related to law enforcement.[197]

Agencies may best deal with gaming concerns by adopting algorithms that are relatively immune to manipulation. For example, the Arnold Foundation claims that PSA-Court, which relies only on objective, verifiable facts concerning a defendant's history, produces risk assessments that are just as accurate as algorithms that rely on subjective statements made by defendants.[198] Azavea has introduced randomness into its HunchLab predictive policing algorithm, which among other things would frustrate efforts to derive patrolling plans even from a disclosed algorithm.[199]

Another concern officials might have is that they don't want to expose their tentative thinking about predictive algorithms. Both FOIA and many state open records acts include an exemption to protect the deliberative process within the executive branch.[200] None of our open records requests was rejected under an executive-branch deliberative process exemption, and so the application of such an exemption to algorithmic processes remains speculative. The deliberative process privilege assumes

---

[196] The City of Cocoa, Florida sent us a document noting that detail about PredPol would not be provided in a public document because "information revealing surveillance techniques, procedures or personnel" is exempt from disclosure under Florida open records law. *See* Fla. Stat. § 119.071(d) ("Any information revealing surveillance techniques or procedures or personnel is exempt from s. 119.07(1) and s. 24(a), Art. I of the State Constitution."). Even if a system for deploying police personnel in particular areas at particular times is a surveillance technique or procedure, a specific exemption for surveillance is not common in open records acts.

[197] As we mentioned above, *see supra* p. 16, government officials also may worry about incidental, detrimental behavioral effects of publicizing algorithms, such as the avoidance of needed mental health treatment by people who learn that having received such treatment is a factor in child welfare risk assessment. Like gaming, this can be a legitimate concern in tension with transparency; there is no open records exemption that addresses it.

[198] *See* Laura & John Arnold Foundation, Developing a National Model for Pretrial Risk Assessment, *available at* http://www.arnoldfoundation.org/wp-content/uploads/2014/02/LJAF-research-summary_PSA-Court_4_1.pdf (noting that other risk assessment instruments "rel[ied] on data that [could] only be gathered through defendant interviews" and that PSA-Court uses only data that is "drawn from the defendant's criminal history.").

[199] *See* A Citizen's Guide to HunchLab, *supra* n. 70, at 10-11.

[200] *See, e.g.*, 5 U.S.C. § 552(b)(5) (exempting "inter-agency or intra-agency memorandums or letters which would not be available by law to a party other than an agency in litigation with the agency"); New York Public Officers Law 87(2)(g) (exempting most "inter-agency or intra-agency materials"); 5 Ill. Comp. Stat. 140/7(f) (exempting "[p]reliminary drafts, notes, recommendations, memoranda and other records in which opinions are expressed, or policies or actions are formulated").

that agencies have already announced a rule and explained its rationale. The point of exempting deliberate process is "to protect against confusing the issues and misleading the public by dissemination of documents suggesting reasons and rationales for a course of action which were not in fact the ultimate reasons for the agency's action."[201] If the government never explains the "rules" of an algorithm or why it was adopted, then there is no authoritative utterance to safeguard from stray deliberation. Indeed, the records created during formulation of the algorithm would be the only window into the rules and rationales bound up in the algorithmic process.

Open records laws often have limited application to the judicial branch.[202] A number of our open records act requests were rejected on the ground that courts were not properly subject to the request. We cannot say this was wrong, but it often should be. The formulation and adoption of an algorithm for a court system bears little resemblance to judicial decisionmaking in individual cases (usually illuminated by public explanation anyway). It is more analogous to the drafting and adoption of a rule of evidence that will be applied to a large set of cases. Judicial rulemaking, like administrative rulemaking, is typically carried out in public. Federal law requires rules promulgated by any federal court other than the Supreme Court "to be prescribed only after giving appropriate public notice and an opportunity for comment,"[203] and the Supreme Court also uses notice-and-comment rulemaking, under procedures issued by the Judicial Conference.[204] State courts have similar public procedures.[205] In the absence of an open records mandate to provide records of the process by which an algorithm was formulated and adopted, courts should consider some form of public process similar to that which they use to adopt and amend rules.

Finally, governments may be worried that some constituents are uncomfortable with the deployment of algorithms, will discern discrimination or unfairness where there is none, or will unduly contest algorithmic recommendations. To avoid what they see as unwarranted controversy, based on distortions or unscientific conclusions or mistakes, governments might rather not publicize algorithmic models. We know of no open records act exemption that prevents controversial matters from disclosure, and while government officials may justifiably fear distortions and unscientific conclusions, controversy is unavoidable in the democratic process. It is often at the heart of it.

---

[201] Coastal States Gas Corp. v. Department of Energy, 617 F.2d 854, 866 (D.C. Cir. 1980).

[202] *See, e.g.*, 5 U.S.C. § 552(f)(1) (defining "agency" to exclude courts); 65 Pa. Stat. § 66.304 (requiring "Judicial agencies" only to provide access to financial records).

[203] 28 U.S.C. § 2071(b).

[204] *See* § 440.20.40 of the Procedures for the Judicial Conference's Committee on Rules of Practice and Procedure and Its Advisory Rules Committees.

[205] Rule 3(a)(1) ("Rulemaking Procedures, Purpose and Applicability"), Ill. Sup. Ct. Rules, as amended June 22, 2017 (providing for a rulemaking process with such elements as "a public record of all . . . proposed rules and proposed amendments" and "an opportunity for comments and suggestions by the public, the bench, and the bar").

IV. FIXES

How can governments promote transparency in their use of predictive algorithms? Legislatures are unlikely to withdraw protection for trade secrets and other confidential information.[206] Even if that were to happen, removal of trade secret protection would not itself solve the problem of inadequate documentation and government possession of records. A more fruitful course will be for governments to use their contracting powers to insist on appropriate record creation, provision, and disclosure.[207] We will first consider provision and disclosure requirements, and then turn to best practices concerning record creation.

A. CONTRACT LANGUAGE REQUIRING PROVISION AND PERMITTING DISCLOSURE OF RECORDS

The agreements between public agencies and contractors that we obtained through open records requests demonstrate that governments do not, and need not, uniformly accede to contractor wishes for nondisclosure and data ownership.

For example, it appears that when the Arnold Foundation drafted a standard Memorandum of Understanding for its PSA program, it included strong, broad language concerning nondisclosure. Courts that did not request changes to that language promised to keep all information they had about the PSA confidential. The Seventh Judicial Circuit of Florida, however, evidently asked for language that provided for significantly narrower nondisclosure duties. It placed the burden on the Arnold Foundation of designating trade secrets, redacting unprotected material, and delivering marked copies to the government. That approach – placing the burden on the contractor to identify and mark specific passages in a document as trade secrets – goes a long way towards avoiding overclaiming trade secrets, and forces the contractor to consider exactly why and how the disclosure of particular information would

[206] For an argument that trade secrecy should not be used to withhold information about a predictive algorithm from a criminal defendant, see Rebecca Wexler, Life, Liberty and Trade Secrets: Intellectual Property in the Criminal Justice System (February 20, 2017), https://ssrn.com/abstract=2920883.

[207] *Cf.*, Joel R. Reidenberg, Lex informatica: The Formulation of Information Policy Rules Through Technology, 76 Texas L. Rev. 553, 589-590 (1998) (arguing for the use of public procurement standards to pursue policy goals).

undermine its competitive position.[208] Such language dovetails with appropriately narrow construction of trade secret exemptions in open records acts.[209]

It is important to recognize that the demand for much narrower nondisclosure language did not cause the Arnold Foundation to refuse to contract with the Seventh Judicial Circuit. The Foundation acceded to the less favorable language even though it provides the PSA for free, and the Seventh Judicial Circuit did not have the bargaining leverage of withholding payment. Nonprofits and foundations need clients just as for-profit companies do – they need to show their donors that they are providing services that are making a difference, and having an impact on how governments run. Thus, governments must understand that they have leverage even if they are not paying for services.[210]

If governments are paying for services, they have additional leverage over nondisclosure and ownership issues. Thus, for example, Illinois' contract to pay Eckerd Kids for the Rapid Safety Feedback service apparently used standard public contracting language containing disclosure and ownership provisions favorable to the State. With regard to disclosure, the contract provides that the default assumption is that all information that Eckerd provides is public[211] – although it could go even further, as the Seventh Judicial Circuit agreement with the Arnold Foundation did, and place the burden on the contractor to make specific, marked claims of trade secrecy or lose the power to object to disclosure. With regard to ownership, the contract provides that Illinois owns everything produced under the contract, including all intellectual property rights in those products.[212] By contrast, when the Alaska Department of Health and Social Services signed an MOU under which Eckerd Kids agreed to provide RSF services without compensation, Alaska promised to treat all Eckerd creations and products as confidential information, and agreed that Eckerd owned everything related to the

---

[208] Similarly, the New York State Education Department contract with American Institutes of Research for the Value Added Measurement project provides that "the contractor shall clearly identify . . . proprietary information [regarding methodologies or measures that are the property of the contractor at the time the contract . . . is executed] and give . . . a license to NYSED to continue using such proprietary information solely for NYSED's educational purposes for a period of ten years from the date of termination of this contract." *See* Contract No. C010834, between the People of the State of New York and American Institutes of Research, dated September 19, 2011, available at [post contract].

[209] *See supra* p. 44.

[210] In some cases, government officials may welcome nondisclosure language, because they want to avoid public scrutiny of their actions. It may be more difficult to deal with a government agency that promises nondisclosure to a contractor so that it has a justification for keeping its own decisionmaking process secret, but in an appropriate case, legal action could be brought challenging such an action as inconsistent with the agency's open government obligations.

[211] *See* State of Illinois Contract Department of Child and Family Services, Rapid Safety Feedback Program, Contract # 5445089016, p. 11, http://www.robertbrauneis.net/algorithms/ILERSFTY16.pdf (hereinafter "IL ERSF Contract").

[212] *See id.* at 11-12.

Rapid Safety Feedback program, including all software and all reports that the software produced.[213]

A contractor that has developed an algorithm intended for multiple jurisdictions without modification will not want to transfer ownership of the source code implementing that algorithm to one jurisdiction. However, if the contractor is providing a custom algorithm for a jurisdiction, then it could be appropriate for that jurisdiction to insist on ownership, or at least a license for its own use and use by other jurisdictions. Thus, for example, Allegheny County's contract with the Auckland Consortium grants a license to the state and federal government to use the software produced under the contract and to authorize others to use it, and grants the county the right to use and distribute anything produced under the contract that is protected by any intellectual property rights.[214] In all cases, government agencies should assert ownership over reports that assess risks in that jurisdiction based on data provided by that jurisdiction. The Illinois contract makes such an assertion,[215] while the Alaska agreement cedes ownership of all reports to Eckerd.[216]

Even very favorable language providing for ownership and disclosure, however, is not effective if no documentation has been created, or if it has never been provided to the government client. Because of the disclosure provisions in the Seventh Judicial Circuit agreement with the Arnold Foundation, that court was able to provide to us information about the PSA risk scales – the percentages of people released pretrial who failed to appear by risk score, both in the original training set and in a validation study – that no other court nor the Arnold Foundation itself would provide. Yet it only was able to provide that information because it happened to be included in a slide presentation made by an Arnold Foundation associate to the court, thus leaving it entirely up to the Arnold Foundation to determine disclosure policy. Accountable governments should make these decisions and link disclosure provisions to demands that records be produced to government, and created if they do not already exist.

B. Creating Records for Accountability

Governments should consciously generate – or demand that their vendors generate – records that will further public understanding of algorithmic processes. This seems to be what is contemplated by the European Union General Data Protection Regulation

---

[213] *See supra* n. 126.

[214] *See* AUT Enterprises Ltd. Contract 9-1-14 to 6-30-15, pp. 37-39, 45-46, http://robertbrauneis.net/algorithms/AlleghenyAUTContract.pdf.

[215] *See* IL ERSF Contract, *supra* n. 235, p. S4, Section 4.8 S b.

[216] *See supra* n. xxx.

(coming into force in 2018), which stipulates that the function of an algorithm must be made understandable to the public.[217]

Ideally, relevant stakeholders would produce a set of best practices for documenting the creation and implementation of predictive algorithms. Such a best practices document could draw on a number of existing models. For example, the Transparency and Accountability Initiative has released a guide to best practices in government transparency, accountability, and civic engagement.[218] The National Federation of Municipal Analysts has promulgated a series of best disclosure practices in connection with the issuance of municipal debt.[219] The Online Trust Alliance has released a number of best practices documents, including the Internet of Things Trust Framework 2.0, a set of privacy and security principles focused on connected home and wearable technologies.[220] Perhaps of most relevance, although at a very high level of abstraction, the U.S. Public Policy Council of the Association for Computing Machinery has produced a set of seven "Principles for Algorithmic Transparency and Accountability."[221]

Although we cannot hope here to provide the kind of best practices statement that would be produced by sustained multi-stakeholder deliberation, we identify based on our research desirable documentation in eight categories: the algorithmic model's general predictive goal; relevant, available, and collectable data; considered exclusion

---

[217] The European Union, as part of its Data Protection Directive, has also given its citizens a right to an explanation of algorithmic decisions (public and private) that "significantly affect" individuals. REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) 2016. *Cf.* Sandra Wachter et al., Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation (December 28, 2016). Available at SSRN: https://ssrn.com/abstract=2903469 (arguing that the Directive "does not, in its current form, implement a right to explanation, but rather a limited 'right to be informed'" of automated decisionmaking); Goodman & Flaxman, supra note xx (identifying developer secrecy, public technical illiteracy, and algorithmic design as barriers to explanation).

[218] *See* Transparency & Accountability Initiative, Opening Government: A guide to best practice in transparency, accountability and civic engagement across the public sector, available at http://www.transparency-initiative.org/wp-content/uploads/2011/07/Opening-Government2.pdf.

[219] *See* National Federation of Municipal Analysts, Disclosure Guidelines, available at http://www.nfma.org/disclosure-guidelines.

[220] *See* Online Trust Alliance, IoT Trust Framework v. 2.0, available at http://otalliance.actonsoftware.com/acton/attachment/6361/f-008d/1/-/-/-/-/IoT%20Trust%20Framework.pdf.

[221] *See* Association for Computing Machinery, U.S. Public Policy Council, Statement on Algorithmic Transparency and Accountability (January 12, 2017), available at http://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf (hereinafter "ACM Algorithmic Transparency and Accountability Statement").

of data; specific predictive criteria; analytic techniques used; principal policy choices made; results of validation studies and audits; and explanation of the predictive algorithm and the algorithm output.

1. General Predictive Goal

Governments should be expected to articulate their goals in using a predictive algorithm. This will provide an important benchmark against which specific criteria can be measured, and may lead to a better understanding of the decisions that algorithmic predictions inform. The goal is not always self-explanatory. For example, the most general goal of an algorithm like PredPol or HunchLab is to predict where and when crimes will occur. Yet a local police force may really be interested in making decisions about where its limited number of patrol officers can most effectively deter crimes, acknowledging that crimes that take place indoors are difficult to deter by patrol. Therefore, the department would more accurately describe its goal as predicting where and when the presence of police patrols would deter crimes.

As part of formulating a general predictive goal, a government may want to take one step further back and articulate the problem it is trying to address. For example, a government that is seeking assistance in predicting which prisoners are most likely to commit crimes if released on parole may be motivated by a variety of concerns. It may want to reduce the prison population because of overcrowding; or it may want to reduce the number of parolees who commit new crimes; or it may be facing challenges about the fairness of its parole decision practices. Each of those situations will likely call for different sensitivities in creating predictive algorithms.

2. Data: Relevant, Available, Collectable

With a predictive goal in mind, the next step is to consider what data could be relevant to making that prediction. It is helpful both for evaluation of an algorithm, and for inducing deliberation, to document what data initially might be thought of as conceivably relevant to predicting the outcome in question. For example, did the data scientists who might have settled on data about a defendant's prior arrest history and employment record also consider data about a defendant's exercise regime and educational background? If not, why not? Most predictive algorithms will be trained on data that has already been collected for some other purpose. Thus, data scientists will go on a search for existing data sources, and it will be important to document where they looked and what they found.

3. Data Exclusion

Data that is available may in the end be excluded from the set of data that is used to train an algorithm, and that will eventually be used as the input to generate a prediction about a particular subject. There are at least five groups of reasons for excluding data: quality concerns, susceptibility to manipulation, time and place limitations, lack of relevance, and policy considerations other than lack of relevance.

a. *Data Quality.* Data scientists may be worried that datasets, or certain data fields, have too many inaccuracies, were not defined consistently as data was collected, or have become corrupted in various ways. For example, addresses may have been manually transcribed from handwritten originals and test as invalid.[222] Or two types of data may have for some period been entered into a single field. Documentation of those issues, and decisions made as to whether to keep the data even with its imperfections, or to exclude it, can be important to assessing the quality of the algorithm produced.

b. *Manipulation and Gaming.* Creators of predictive algorithms may also decide to exclude some types of data because it is subject to manipulation or "gaming," and thus undermines either the accuracy of the training data, or the accuracy of the input to the completed algorithm. For example, as mentioned above, the Arnold Foundation decided to create a pretrial release algorithm that would not require as input any facts gathered in an interview with the criminal defendant.[223] This exclusion was partly motivated by the concern that information collected during an interview, when the defendant knows that the responses can determine pre-trial release, is subject to manipulation.

c. *Time and Place Limitations.* Data is necessarily collected about subjects who are acting in different times and places. All other things being equal, the larger the training dataset, the better. But all other things may not be equal. The risk of recidivism ten years ago may be different today for prisoners with the same profile, due to the economy, available social services, and many other factors. If data subsets from different years exhibit markedly different correlations, a decision may be made to exclude older data as stale. On the other hand, if the goal is to predict whether a parolee will commit a crime in the next five years, then the training dataset must exclude data about prisoners who have been paroled less than five years ago, because newer parolees will not have a sufficiently long track record. In some instances, then, some data may have to be excluded as too old, and other data as too new.[224]

In the case of HunchLab, the Lincoln Police Department revealed that it uses HunchLab once a day, and that each day it inputs police incident reports for the

---

[222] *Cf.* Julia Andre, Luis Ceferino & Thomas Trinelle, Prediction algorithm for crime recidivism, available at http://cs229.stanford.edu/proj2015/250_report.pdf (cautioning that "publicly available datasets [of recidivism of released inmates] are ancient, due to prescriptions, which means that they are often number re-transcription of manually stored data").

[223] *See* Laura & John Arnold Foundation, Developing a National Model for Pretrial Risk Assessment, *available at* http://www.arnoldfoundation.org/wp-content/uploads/2014/02/LJAF-research-summary_PSA-Court_4_1.pdf.

[224] On the choice of time and place limitations for data, see Andreas M. Olligschlaeger, Crime Forecasting on a Shoestring Budget, Crime Mapping & Analysis News 8, 9-10 (Spring 2015), https://crimemapping.info/wp-content/uploads/2015/03/CrimeMappingNews_Issue23.pdf.

previous 30 days.[225] The choice of a 30-day window obviously involves a balance of competing factors. Restricting input to the past month keeps the data relatively fresh, and allows for inquiry into weekly and monthly cycles of activity. At the same time, it does not allow for inquiry into seasonal cycles, and may lead to very thin data on relatively uncommon types of crimes.

Algorithm developers must also make judgments about the geographic scope of training and input data. Due to different social and economic conditions, and perhaps more controversially due to different ethnic composition, income profile, or other factors, a group of defendants from one area – perhaps an urban area – who are otherwise similar to a group of defendants from a second area – perhaps a rural area – may pose different risks of pretrial flight.

We know that the Arnold Public Safety Assessment algorithm was trained on data that was aggregated from 300 different jurisdictions nationwide.[226] We do not know if the Arnold Foundation tested whether subsets of that dataset from different states or regions exhibited the same predictive correlations as the dataset as a whole. If data from different regions exhibit substantially different predictive correlations, a decision may be made to geographically restrict the dataset. Whether or not the dataset is restricted by time and place, it may be a best practice to test for difference across time and place and document the results.

d. Relevance. Some data elements may be excluded because they do not seem to be sufficiently correlated with the outcome sought to be predicted. It would be useful to document that exclusion, and the threshold of predictive value below which the excluded data fell.

e. Policy Reasons Other Than Relevance. Perhaps most notably and controversially, certain data will be excluded in spite of its potential predictive value, for a variety of policy reasons. For example, the Arnold Foundation promotes as an advantage of its algorithm that it does not take into account matters such as "race, gender, income, education, home address, drug use history, family status, marital status, national origin, employment, [or] religion."[227] Immutable characteristics such as race and gender are constitutionally problematic; home address may in many cases be closely correlated with race. The decision to exclude characteristics such as level of education and drug

---

[225] *See* Letter of November 30, 2016 of Tonya Peters, available at https://www.muckrock.com/foi/lincoln-4033/lincoln-police-department-hunchlab-documents-30110/#file-110327.

[226] *See* The Laura and John Arnold Foundation, Developing a National Model for Pretrial Risk Assessment, at 3, available at http://www.arnoldfoundation.org/wp-content/uploads/2014/02/LJAF-research-summary_PSA-Court_4_1.pdf.

[227] *See* Arnold Foundation, the Public Safety Assessment (PSA), *available at* http://www.arnoldfoundation.org/wp-content/uploads/PSA-Infographic.pdf (citing a Department of Justice study that 52% of violent crimes were not reported to police between 2006 and 2010).

use history, if they are found to have substantial predictive value, would presumably be more controversial, and should be documented.

4. Specific Predictive Criteria

We noted above that it can be useful to articulate a general predictive goal that an algorithm development project will pursue. Once decisions have been made about what training data to use, however, it will likely turn out that the actual predictions will have to be described somewhat differently than the original predictive goal. For example, the general predictive goal of an algorithm may be to predict where and when crime will occur, most likely using reported crimes as the training and input data. Because crime is underreported,[228] this data will not optimally support the general goal.

More importantly, however, crimes are reported at different rates in different neighborhoods.[229] For example, one study found that simple assaults were less likely to be reported in disadvantaged neighborhoods.[230] Another found that crimes were particularly underreported in heavily immigrant neighborhoods.[231] A third found that reporting of crimes tends to increase with the age of the victim, so that neighborhoods with older residents will likely report a higher percentage of crimes.[232]

These issues are not limited to predictive policing. For example, Allegheny County was most interested in predicting when reported child maltreatment was likely to result in serious injury or death, but it decided that it could not build an algorithm that would do so directly, because the cases in which serious injury or death actually occurred provided (thankfully) too few data points. It therefore decided instead to use the proxies of placement in a foster home and additional reports of maltreatment, for

---

[228] *See* United States Department of Justice, Bureau of Justice Statistics, Victimizations Not Reported to Police, 2006-2010 (August 2012), available at https://www.bjs.gov/content/pub/pdf/vnrp0610.pdf.

[229] On the general divergence of reported crime from true crime rates, see David Robinson & Logan Koepke, Stuck in a Pattern: Early evidence on predictive policing and civil rights (2016), p. 5, https://www.teamupturn.com/static/reports/2016/predictive-policing/files/Upturn_-_Stuck_In_a_Pattern_v.1.01.pdf.

[230] *See* Eric P. Baumer, Neighborhood Disadvantage and Police Notification by Victims of Violence, 40 Criminology 579 (2002).

[231] *See* Carmen M. Gutierrez & David S. Kirk, Silence Speaks: The Relationship Between Immigration and the Underreporting of Crime, Crime & Delinquency (September 2015)

[232] *See* Stacey J. Bosick, Callie Marie Rennison, Angela R. Gover & Mary Dodge, Reporting Violence to the Police: Predictors Through the Life Course, 40 J. Criminal Justice 441 (2012). Admirably, Azavea, Inc., the creator of HunchLab, discusses in some detail its choice of reported crimes as training data, the reasons why it has made that choice, and the type of crime reports it prefers. *See* A Citizen's Guide to HunchLab 2, 25-26 (draft July 11, 2017), http://www.robertbrauneis.net/algorithms/HunchLabACitizensGuide.pdf.

reasons that it explains at length in the Auckland Consortium report.[233] Similarly, the COMPAS recidivism algorithm is trained on data about repeat arrests for crimes, not data about convictions;[234] although the Arnold Foundation has not disclosed details about its PSA training data, it almost certainly also uses arrests rather than convictions. It is important to understand how those two may diverge. Abe Gong asks us to consider, "What if police officers are more likely to pursue, search and arrest black suspects than white suspects? What if law enforcement deploys a disproportionate amount of force or uses more aggressive policing tactics in black neighborhoods?"[235] Arrests of minority community members will be skewed artificially high.

5. Analytic and Development Techniques Used

A relatively small number of analytic techniques are used to discover correlations between characteristics or features of subjects of prediction. Among the most popular are regression techniques (linear, logistic, and polynomial), random forests, neural networks, and support vector machines.[236] It is helpful to document which techniques were tried, and which chosen and why. For example, linear regression may be appropriate when it is thought likely that there is indeed a linear relationship between one or more inputs and the output – for example, between the age of a defendant and the likelihood that the defendant will commit a crime if released before trial. The result of performing linear regression may be some line with statistical significance, but that doesn't mean it fits the data better than another analytic technique, which produces a different predictive model that is non-linear (e.g., because it might use cutoffs of particular ages).

There are also standard algorithm development techniques in use, such as dividing a dataset randomly into subsets that will be used for training an algorithm, and then testing it ("validation") in one or more stages.[237] Documentation of those development techniques is also likely a best practice.

---

[233] *See* Rhema Vaithianathan, Emily Putnam-Hornstein, Nan Jiang, Parma Nand & Tim Maloney, Developing Predictive Models to Support Child Maltreatment Hotline Screening Decisions: Allegheny County Methodology and Implementation 9-11 (2017), http://www.alleghenycountyanalytics.us/wp-content/uploads/2017/04/Developing-Predictive-Risk-Models-package-with-cover-1-to-post-1.pdf.

[234] *See* Northpointe, COMPAS Risk & Need Assessment System, p. 2 http://www.northpointeinc.com/files/downloads/FAQ_Document.pdf (the General Recidivism Risk Scale algorithm is trained on data on whether defendants have been arrested within two years of an intake assessment).

[235] Abe Gong, Ethics of Powerful Algorithms (2 of 4), https://medium.com/@AbeGong/ethics-for-powerful-algorithms-2-of-3-5bf750ce4c54.

[236] *See, e.g.*, Shai Shalev-Schwartz & Shai Ben-David, Understanding Machine Learning: From Theory to Algorithms 89-240 (2014).

[237] *See, e.g.*, Yaser S. Abu-Mostafa, Malik Magdon-Ismail & Hsuan-Tien Lin, Learning from Data: A Short Course 138-154 (2012).

6. Principal Policy Choices

We have mentioned a number of different types of policy choices made in the development of an algorithm. One is the decision to exclude otherwise relevant data for various reasons. Another is the decision to weight false negatives and false positives equally or differently. Those choices should be documented, along with accounts of why they were made the way they were.

7. Validation Studies, Audits, Logging, and Nontransparent Accountability

Pre-implementation validation is a standard step in the initial development of a predictive algorithm. However, after an algorithm has been put into service, additional post-implementation validation studies may be conducted regarding the predictive strength of the algorithm, and any output biases that it may be producing, under real-world conditions. Best practices could be developed about when and how such studies should be conducted, and when it is appropriate to insist that the studies be conducted by an independent entity. Public clients could require that such studies be conducted on their cases and delivered to them.

An alternative or addition to a validation study is an audit. Where optimal disclosure won't happen, for trade secret, security, or privacy reasons, it would be important to have a third-party confidential audit of algorithm development.[238] Public clients could insist on an audit whenever an algorithm misses certain targets, or when the clients discover evidence that the development process was flawed. It would also be appropriate to require the developer to keep a log containing many or all of the categories of documentation described above, even though the complete log would not ordinarily be disclosed, just in case an audit became necessary.[239] Public entities should also contract for audits of algorithm implementation, which is what the Seventh Judicial Circuit got of its implementation of the PSA algorithm (performed by an Arnold Foundation subcontractor). Public clients should know and be able to reveal to the public whether they are inputting data and interpreting results correctly.

8. Algorithm and Output Explanations

It will often be important to provide a plain-language explanation of the correlations upon which an algorithm is based, and of the general path that it takes to its prediction, whether that be a formula that weights factors, a decision tree, or some other path.[240]

---

[238] On algorithm audits, see Christian Sandvig, Kevin Hamilton, Karrie Karahalios & Cedric Langbort, Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms, http://tiny.cc/61wrmy.

[239] *See* ACM Algorithmic Transparency and Accountability Statement, *supra* note 221, Principle 6 ("Auditability: Models, algorithms, data, and decisions should be recorded so that they can be audited in cases where harm is suspected.").

[240] *See id.*, Principle 4 ("Explanation: Systems and institutions that use algorithmic decision-making are encouraged to produce explanations regarding both the procedures followed by the

If the algorithm is so complicated that a plain-language explanation does not seem possible, that should probably disclosed as well, so that those who are using the predictive output of the algorithm understand that it is a black box, unconnected to any articulable explanation or causal theory. If an interpretable algorithm performs as well as a non-interpretable algorithm, governments should prefer the interpretable one for the sake of government capacity as well as public transparency. If the line agents (or people they trust) understand the algorithm, they will be better equipped to accept its judgment or override it.[241]

It will also often be important to provide explanations of the algorithm's output. That is particularly true when the algorithm produces an uncalibrated scale, like the PSA's risk scales of one to six. In a validation study conducted on early implementation of the algorithm, almost nine out of ten defendants who earned the lowest score for risk of pre-trial flight actually did appear at trial; for those who earned the *highest* risk score, seven out of ten appeared. If pretrial services officials and judges are not aware of those percentages, they might assume that the difference between the lowest and highest risk scores is greater than it actually is, or they may have different assumptions about how low a risk a "low-risk" defendant poses, or how high a risk a "high-risk" defendant poses.[242]

V. CONCLUSION

There will always be value for public entities to use open source code, or to otherwise release the code running predictive analytics. But access to code will not usually be necessary to achieve meaningful transparency and sometimes will not even help. What public entities should be more focused on is undertaking the design, procurement, and implementation of algorithmic processes in more thoughtful and transparent ways. Public entity contracts should require the vendors to create and deliver records that explain key policy decisions and validation efforts, without necessarily disclosing precise formulas or algorithms. Those records can then be released and support open policy debates without adversely affecting the contractor's competitive position. To the extent that irreducible trade secrets remain in predictive algorithm projects, government records custodians responding to open records requests should construe

---

algorithm and the specific decisions that are made. This is particularly important in public policy contexts."), Diakopoulos, Algorithmic accountability, *supra* note 5, at 411 (recommending that a transparency policy for algorithms include "the definitions, operationalizations, or thresholds used by similarity or classification algorithms").

[241] On the development of interpretable algorithms, *see* Jiaming Zeng, Berk Ustun & Cynthia Rudin, Interpretable Classification Models for Recidivism Prediction (2016), available at https://arxiv.org/abs/1503.07810

[242] Assessing whether subjects are grouped in a way that reflects risk differences is referred to as "calibration." *See, e.g.*, Nicholas Serrano, *Calibration Strategies to Validate Predictive Models: Is New Always Better?*, 38 Intensive Care Medicine 1246 (2012), https://link.springer.com/article/10.1007/s00134-012-2579-z.

those claims narrowly. Courts should do the same, requiring contractors to release records (even in redacted form) that will not weaken their competitive position.