Algorithmic Transparency for the Smart City

Robert Brauneis & Ellen P. Goodman[1]

20 YALE J. L. & TECH. 103 (2018)

> *"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."*[2]

As artificial intelligence and big data analytics increasingly replace human decision making, questions about algorithmic ethics become more pressing. Many are concerned that an algorithmic society is too opaque to be accountable for its behavior. An individual can be denied parole or credit, fired, or not hired for reasons that she will never know and which cannot be articulated. In the public sector, the opacity of algorithmic decision making is particularly problematic, both because governmental decisions may be especially weighty and because democratically elected governments have special duties of accountability.

We set out to test the limits of transparency around governmental deployment of big data analytics, contributing to the literature on algorithmic accountability with a thorough study of the opacity of governmental predictive algorithms. Using open records processes, we focused our investigation on local and state government deployment 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 data analytics to private vendors, which may make design and policy choices unseen by client agencies, the public, or both. To test how impenetrable the resulting "black box" algorithms are, we filed forty-two open

---

[1] 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 Eighteenth Annual Congress of the European Intellectual Property Institutes Network.

[2] Peter Stone et al., *Overview*, ONE HUNDRED YEAR STUDY ON ARTIFICIAL INTELLIGENCE (AI100) (Aug. 1, 2016), http://ai100.stanford.edu/2016-report/overview [http://perma.cc/3GQY-HRWV].

records requests in twenty-three 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 partnerships. The specific goal was to assess whether open records processes would enable citizens to discover what policy judgments these algorithms embody and to evaluate their utility and fairness.

To do this work, we identified what meaningful "algorithmic transparency" entails. We found that in almost every case, it was not provided. Over-broad assertions of trade secrecy were a problem. But, contrary to conventional wisdom, trade secrets properly understood are not the biggest obstacle, as release of the trade-secret-protected code used to execute predictive models will not usually be necessary for meaningful 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.

TABLE OF CONTENTS

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.[3] 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.[4] 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 areas such as criminal justice, food safety, social services, and transportation.[5]

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.[6] These

---

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

4 *See* ANDREW GUTHRIE FERGUSON, THE RISE OF BIG DATA POLICING: SURVEILLANCE, RACE, AND THE FUTURE OF LAW ENFORCEMENT (2017); Andrew Guthrie Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1115 (2017); Elizabeth E. Joh, *The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing*, 10 HARV. L. & POL'Y REV. 15, 38 (2016).

5 On smart-city algorithms generally, see *infra* text accompanying notes 27-36; on the algorithms about which we filed open records requests, see text accompanying notes 123-183.

6 *See, e.g.*, ROB KITCHIN, THE DATA REVOLUTION: BIG DATA, OPEN DATA, DATA INFRASTRUCTURES AND THEIR CONSEQUENCES (2014); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION (2016); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015); Michael Ananny, *Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness*, 41 SCI. TECH. & HUM. VALUES 93 (2015); David Beer, *The Social Power of Algorithms*, 20 J. INFO. COMM. & SOC. 1 (2016); Taina Bucher, *'Want To Be on the Top?' Algorithmic Power and the Threat of Invisibility on Facebook*, 14 NEW MEDIA & SOC'Y 1164 (2014); Jenna Burrell, *How the Machine 'Thinks:' Understanding Opacity in Machine Learning Algorithms*, 3 BIG DATA & SOC'Y 1 (2016); Danielle Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014); Kate Crawford, *Can an Algorithm be Agonistic? Ten Scenes from Life in Calculated Publics*, 41 SCI. TECH. & HUM. VALUES 77 (2016); Nicholas Diakopoulos, *Algorithmic Accountability: Journalistic Investigation of Computational Power Structures*, 3 DIGITAL JOURNALISM 398 (2015); Tarleton Gillespie, *The Relevance of Algorithms*, *in* MEDIA TECHNOLOGIES: ESSAYS ON COMMUNICATION, MATERIALITY, AND SOCIETY 167 (Tarleton Gillespie et al. ed., 2014); Rob

black boxes are impervious to question, and many worry that they may be discriminatory,[7] erroneous, or otherwise problematic.[8] 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.[9]

Commentators have called for more transparency across all implementations of artificial intelligence.[10] There are special

---

Kitchin, *Thinking Critically About and Researching Algorithms*, 20 INFO. COMM. & SOC'Y 14 (2016) [hereinafter Kitchin, *Thinking Critically*]; Christian Sandvig, *Seeing the Sort: The Aesthetic and Industrial Defense of "The Algorithm"*, 12 J. NEW MEDIA CAUCUS, 1 (2015); Christian Sandvig et al., *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms* (2014), http://www.personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20-%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf [http://perma.cc/M9FK-3R2V]; Zynep Tufekci, *Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency*, 13 J. ON TELECOM. & HIGH TECH. L. 203 (2015); Malte Ziewitz, *Governing Algorithms: Myth, Mess, and Methods*, 41 SCI. TECH. & HUM. VALUES 3 (2015). *See generally* Tarleton Gillespie & Nick Seaver, *Critical Algorithm Studies: A Reading List*, SOC. MEDIA COLLECTIVE RES. BLOG http://socialmediacollective.org/reading-lists/critical-algorithm-studies/ [http://perma.cc/TBF5-DEY4] (compiling sources).

7 *See, e.g.*, FED. TRADE COMM'N, BIG DATA: A TOOL FOR INCLUSION OR EXCLUSION? UNDERSTANDING THE ISSUES (Jan. 2016), http://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report [http://perma.cc/B922-N6U4]; Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671 (2016); Tufecki, *supra* note 6; Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), http://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [http://perma.cc/684J-45ZP] (evaluating algorithmic risk assessments used by judges to set bail amounts in Fort Lauderdale, Florida).

8 *See generally* Lee Rainie & Janna Anderson, *Code-Dependent: Pros and Cons of the Algorithm Age*, PEW REP. (Feb. 9, 2017), http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/ [http://perma.cc/MCQ8-APDG] (summarizing examples of the risks of using algorithms broadly).

9 *See, e.g.*, Nicholas Diakopoulos, *We Need To Know the Algorithms the Government Uses To Make Important Decisions About Us*, CONVERSATION (May 23, 2016), http://theconversation.com/we-need-to-know-the-algorithms-the-government-uses-to-make-important-decisions-about-us-57869 [http://perma.cc/N522-46V4] (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/ [http://perma.cc/RPM7-LV3K] (describing journalists suing the 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).

10 *See, e.g.*, *23 Principles for Beneficial Artificial Intelligence*, FUTURE LIFE INST., (Jan. 17, 2017), http://futureoflife.org/ai-principles/ [http://perma.cc/THU7-

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 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 by identifying some of the most common uses of big data prediction in that sector. We identified algorithms developed by foundations, private corporations, and government entities and those used in criminal justice and in civil applications. We then assembled a "portfolio" of open records requests targeting a variety of uses and jurisdictions. Using MuckRock, the nonprofit collaborative platform for filing open records requests,[11] we filed 42 requests in 23 states for records relating to six predictive algorithms.[12] 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.[13] Software is a "record"

CPGM] (noting that 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).

11 *About Us*, MUCKROCK, http://www.muckrock.com/about/ [http://perma.cc/L8SV-TYDL]. In one case (Allegheny County Child and Family Services), we filed the requests separately, not using the platform.

12 Our project page on MuckRock can be found at *Uncovering Algorithms*, MUCKROCK, http://www.muckrock.com/project/uncovering-algorithms-84/ [http://perma.cc/HNW2-ZSUA]. That page links to 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.

13 *See, e.g.*, FLA. STAT. § 119.011(12) (2017) (providing that a "public record" open for inspection "means all documents, papers, letters, maps, books, tapes,

disclosable under the federal Freedom of Information Act (FOIA),[14] as well as under many state laws, but not all.[15] We sought records including correspondence, contracts, software, training materials, existing and planned validation studies, and other documentation. Although our focus was local and state government, we suspect our findings are generalizable to other governmental entities working with private vendors.

We conclude from the results of those requests and associated investigation 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 which makes them difficult to interpret even when records are revealed. We save this issue for another day.

In Part I of this Article, we introduce basic concepts such as clinical judgment, actuarial judgment, and predictive algorithms; trace the development of smart city algorithmic governance; and survey the promise and perils of governing by algorithm. In Part II, we develop the concept of "meaningful transparency" regarding predictive algorithms. That involves exploring what the public needs to know about the politics embedded in these programs, and about their utility, fairness, and impact on governmental capacity. Part III describes the open records requests we submitted to various jurisdictions about their deployment of predictive algorithms, and the responses we received. Part IV identifies obstacles to greater transparency with respect to algorithmic processes. Part V suggests mitigation techniques to maximize algorithmic transparency, and presents eight principal types of information

---

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").

14 *See generally* Katherine Fink, *Opening the Government's Black Boxes: Freedom of Information and Algorithmic Accountability*, INFO. COMM. & SOC'Y (2017) (discussing research on federal agency responses to FOIA requests for source code).

15 *Compare* ARK. CODE ANN. § 25-19-103(5)(B) (providing that the definition of a "public record" does not include "software acquired by purchase, lease, or license."), *with* FLA. STAT. § 119.011(1) (providing that "[d]ata processing software" is included in the definition of a "public record."). *See generally infra* notes 113-117 (providing examples of state open records law coverage).

that records concerning publicly implemented algorithms should contain. Part VI concludes.

I. THE PROMISE AND PERILS OF ALGORITHMIC GOVERNANCE

A. From Clinical Prediction to Smart City Algorithmic Governance

Algorithmic governance is made possible by vast increases in computing power and networking, which enable the collection, storage, and analysis of large amounts of data. Cities seek to harness that data to rationalize and automate the operation of public services and infrastructure, such as health services, public safety, criminal justice, education, transportation, and energy.[16] The limitations of local government make private contractors central to this process, giving rise to accountability problems characteristic of policy outsourcing. This movement from public need to private technology support has its origins in the preference for "actuarial" over "clinical" prediction.

1. Clinical Versus Actuarial Prediction and Judgment

Government officials who allocate public resources and deploy coercive police power often use what has been called "clinical prediction" to make their decisions.[17] From training, apprenticeship, and experience, those officials develop a sense of how people behave and what consequences an administrative decision is likely to have. A seasoned caseworker has a sense of whether a child is likely to be mistreated in a household. A judge predicts whether a prisoner will commit another crime if paroled. And a public college admissions officer can roughly predict what kind of scholarship offer will prompt an admitted

---

[16] *See* Hafedh Chourabi et al., *Understanding Smart Cities: An Integrative Framework*, *in* PROC. 2012 45TH HAW. INT'L CONF. ON SYS. SCI. (HICSS 2012), http://dx.doi.org/10.1109/HICSS.2012.615 [http://perma.cc/K5GS-66N5] (describing and synthesizing various conceptions of the smart city); Lilian Edwards, *Privacy, Security and Data Protection in Smart Cities: A Critical EU Law Perspective*, 2 EUR. DATA PROTECTION L. REV. 28, (2016); *see also* Nils Walravens & Pieter Ballon, *Platform Business Models for Smart Cities: From Control and Value to Governance and Public Value*, 51 IEEE COMM. MAG. 72 (2013) (discussing the 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).

[17] *See, e.g.*, PAUL E. MEEHL, CLINICAL VERSUS STATISTICAL PREDICTION: A THEORETICAL ANALYSIS AND A REVIEW OF THE EVIDENCE (1954); Theodore R. Sarbin, *A Contribution to the Study of Actuarial and Individual Methods of Prediction*, 48 AM. J. SOC. 593 (1943).

student to enroll. Clinical prediction, or clinical judgment, cannot be fully formalized or articulated, as it is based on applying the accumulated wisdom of an individual to a particular situation, informed by an open-ended inquiry into relevant circumstances.

Clinical prediction stands in contrast to actuarial prediction, which has also been called mechanical, statistical, or algorithmic prediction or judgment.[18] An actuarial approach analyzes data about subjects to discover correlations between features or characteristics and outcomes. It is decidedly not open-ended. Analysis of data about parolees, for example, could hypothetically reveal that those between the ages of twenty and thirty were arrested twice as often while on parole as those between ages fifty and sixty. The data definitions, methods of analysis, and correlations of actuarial prediction can be formalized and articulated, unlike those of clinical prediction. The resulting predictions or judgments, however, are never based on all the circumstances of a particular situation, because actuarial analysis always operates with a finite number of data fields.[19] For example, while the age, sex, and criminal history of a criminal defendant may be considered, the quality of her familial relationships and community connections may not be.

Actuarial judgment has been around for a long time. The first life insurance company to sell policies based on actuarial tables was founded in London in 1762.[20] Ninety years ago, in 1928, Ernest Burgess created a formula for predicting recidivism among parolees, based on statistical analysis of 21 factors[21]—a formula that a later review determined to be more accurate than clinical prognoses of prison psychiatrists.[22] Over sixty years ago, the use of actuarial prediction was widespread enough that Paul Meehl published a famous book contrasting clinical and

---

18 *See* William M. Grove et al., *Clinical Versus Mechanical Prediction: A Meta-Analysis*, 2000 PSYCHOL. ASSESSMENT 19; Jack Sawyer, *Measurement and Prediction, Clinical and Statistical*, 66 PSYCHOL. BULL. 178 (1966).

19 On the correlation between clinical judgment and particularism, and between actuarial judgment and generalization, see FREDERICK SCHAUER, PROFILES, PROBABILITIES, AND STEREOTYPES 19-22 (2003).

20 *See* MAURICE EDWARD OGBORN, EQUITABLE ASSURANCES: THE STORY OF LIFE ASSURANCE IN THE EXPERIENCE OF THE EQUITABLE LIFE ASSURANCE SOCIETY, 1762-1962, at 39 (1962).

21 *See* Ernest W. Burgess, *Factors Determining Success or Failure on Parole*, *in* ANDREW A. BRUCE ET AL., THE WORKINGS OF THE INDETERMINATE SENTENCE LAW AND THE PAROLE SYSTEM IN ILLINOIS 205 (1928); *see also* BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING AND PUNISHING IN AN ACTUARIAL AGE 47-76 (2007) (discussing Burgess's work and its significance); Karl F. Schuessler, *Parole Prediction: Its History and Status*, 45 J. CRIM. L. & CRIMINOLOGY 425 (1955) (discussing the history of statistical prediction in relation to parole decisions, including Burgess's work).

22 *See* Daniel Glaser, *A Reconsideration of some Parole Prediction Factors*, 19 AM. SOC. REV. 335 (1954).

statistical prediction.[23] And of course governments have long been applying, in piecemeal form at least, more formal statistical methods alongside informal clinical judgment, as they consider matters such as the degree to which certain safety improvements would likely reduce traffic deaths, or the likely change in demand for electricity in a community over some period in the future. Yet the application of actuarial judgment in government (as well as in business) has recently picked up tremendous momentum, thanks to the accumulation of large datasets for analysis and advances in computing power and machine learning theory that have enabled much more complex analysis of those datasets.

2. Predictive Algorithms and Machine Learning

In the era of big data, actuarial judgment is implemented through the creation and use of predictive algorithms. Predictive algorithms are constructed through analysis of large datasets to reveal correlations between various features (of a person, circumstance, or activity) and desired or objectionable outcomes.[24] Typically, that analysis is performed with the assistance of machine learning processes, processes in which computers may test thousands or millions of complex correlations to see which best fits the data. Machine learning processes are powerful—they comb through a very large number of possibilities—and relatively objective—the computer has no idea whether a particular variable represents a feature that a person might consider irrelevant, like shoe size, or sensitive, like race, but simply tests the strength of any correlation between that variable and the variable representing the outcome. The set of correlations with the best fit becomes a model that will estimate the likelihood of future behavior or events (the output) when given relevant facts (the input).[25] An algorithmic process will therefore typically involve (1) the construction of a model to achieve some goal, based on analysis of collected historical data; (2) the coding of an algorithm that implements this model; (3) collection of data about subjects to provide inputs for the algorithm; (4) application of the prescribed algorithmic operations on the input data; and (5) outputs in the form of

---

23 *See* MEEHL, *supra* note 17.

24 EXEC. OFFICE OF THE PRESIDENT, BIG DATA: A REPORT ON ALGORITHMIC SYSTEMS, OPPORTUNITY, AND CIVIL RIGHTS, 8 (May 2016) (describing machine learning as the "science of getting computers to act without being explicitly programmed" (quoting Andrew Ng, *Coursera Machine Learning Course*, STAN. U. 2016)).

25 Robin K. Hill, *What an Algorithm Is*, 29 PHIL. & TECH. 35 (2015).

predictions or recommendations based on the chain of data analysis.[26]

3. Algorithmic Governance and Smart Cities

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.[27] Implementation of algorithms at the local level is part of a broader move towards data-driven decision making, and must be understood in the context of the "smart city" agenda.

In the twenty-first century, cities and counties have increasingly turned to "digital hardware and software, producing massive amounts of data about urban processes."[28] At first, the integration of digital technologies into governance involved rudimentary e-government initiatives and digitizing governmental resources.[29] In the past half-decade, local governments have deployed more extensive analytics and begun to exploit sensor networks, ubiquitous communications, and computing.[30] All of this work—the collection, analysis, and use of data—requires technical know-how and infrastructure that most governments lack. Cities are being asked to handle more with fewer resources as they transition to data-based governance. Private technology companies provide "solutions" which can be difficult for city managers to assess.

Local government depends on public-private partnerships to develop the analytics necessary for "smart" urban systems.[31] Controversially, private entities have been at the leading edge of the entire smart city movement.[32] Indeed, IBM registered the

---

26 See Tal Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503, 1517-20; Gillespie, *supra* note 6, at 167; Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 640 n.14 (2017).

27 Marijn Janssen & George Kuk, *The Challenges and Limits of Big Data Algorithms in Technocratic Governance*, GOV'T INFO. Q. 33, 371-77 (2016) (discussing how algorithms and big data become a form of governance, often impervious to interrogation or explanation).

28 Alan Wiig & Elvin Wyly, *Introduction: Thinking Through the Politics of the Smart City*, 37 URB. GEOGRAPHY 485, 488 (2016); *see also* Rob Kitchin, *The Real-Time City? Big Data and Smart Urbanism*, 79 GEOJOURNAL 1 (2014).

29 *See generally* STEPHEN GOLDSMITH & SUSAN CRAWFORD, THE RESPONSIVE CITY: ENGAGING COMMUNITIES THROUGH DATA-SMART GOVERNANCE (2014) (tracing the stages of integration of digital technology into governance).

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

31 *See, e.g.*, Alberto Vanolo, *Smartmentality: The Smart City as Disciplinary Strategy*, 51 URB. STUD. 883 (2013) (describing the centrality of public-private partnership to the smart city vision and implementation).

32 *See* Janine S. Hiller & Jordan M. Blanke, *Smart Cities, Big Data, and the Resilience of Privacy*, 68 HASTINGS L.J. 309 (2017) (describing the corporate framing of smart cities).

phrase "smarter cities" as a trademark as part of its campaign to market technology-driven urban management.[33] Cisco has been similarly active.[34] Sidewalk Labs, a subsidiary of Alphabet (which owns Google), is redeveloping the Toronto waterfront.[35] The governance of this smart waterfront district will be equally split between Sidewalk Labs and the government agency, Waterfront Toronto.[36]

B. Promises and Perils

There is both good and bad in smart city reliance on private technology companies, as well as in governmental deployment of actuarial judgment through algorithmic processes more generally.

On the good side, algorithmically informed decision making promises increased efficacy and fairness in the delivery of government services. As has been demonstrated in medicine, actuarial prediction is sometimes measurably better than clinical prediction: formalized analysis of datasets can result in better assessments of risk than informal professional hunches developed over years of experience in practice.[37] 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.[38] Integrating data across siloed administrative domains, such as

---

33 *See* SMARTER CITIES, Registration No. 4,033,245; Ola Söderström et al., *Smart Cities as Corporate Storytelling*, 18 CITY 307 (2014); *see also* Alan Wiig, *IBM's Smart City as Techno-Utopian Policy Mobility,* 19 CITY 158 (2015) (exploring the utopian rhetoric and extravagant promises of early IBM smart city initiatives); Alan Wiig, *The Empty Rhetoric of the Smart City: From Digital Inclusion to Economic Promotion in Philadelphia*, 37 URB. GEOGRAPHY 535, 540 (2016) (explaining that IBM's Smarter Cities Challenge offered cities partnerships with corporate "consultants and technology specialists [that] will help municipalities analyze and prioritize their needs, review strengths and weaknesses, and learn from the successful strategies used by other cities").

34 Gordon Falconer & Shane Mitchell, *Smart City Framework: A Systematic Process of Enabling Smart + Connected Communities*, CISCO (2012), http://www.cisco.com/c/dam/en_us/about/ac79/docs/ps/motm/Smart-City-Framework.pdf [http://perma.cc/D3ZA-K8NC].

35 *Innovation and Funding Partner Framework Agreement: Summary of Key Terms for Public Disclosure*, SIDEWALK TORONTO (Nov. 1, 2017), http://sidewalktoronto.ca/wp-content/uploads/2017/10/Waterfront-Toronto-Agreement-Summary.pdf [http://perma.cc/6JAL-HEQ4].

36 *Id.*

37 *See, e.g.*, Grove et al., *supra* note 18.

38 *See* Edward L. Glaeser et al., *Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life* (Harv. Bus. Sch. NOM Unit, Working Paper No. 16-065, 2015), http://dash.harvard.edu/bitstream/handle/1/24009688/16-065.pdf?sequence=1 [http://perma.cc/QB2R-CFLG].

education and human services, and then using that data to prioritize families in need of government help, can improve social service delivery.[39]

Algorithmically informed decision making can also help government officials avoid the biases, explicit or implicit, that may creep into less formal, "hunch"-based decision making.[40] 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, 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 the risk of recidivism when fed input information about the prisoner.[41] While constructing an algorithm, government officials can explicitly decide to exclude sensitive attributes such as race, ethnicity or religion, as well as categories of data that serve as proxies for those sensitive attributes, from consideration if they conclude that such consideration would be unfair.

At the same time, predictive algorithms are hardly infallible and pose special risks, especially when substantially controlled by private partners. When improperly developed or implemented, predictive algorithms can turn out to be less accurate than the clinical judgment of government officials, and they can formalize and mask biases embedded in the data on which they are trained. Moreover, as we will discuss below, algorithms may enact policy judgments that diverge from the preferences of the electorate or its elected representatives.

The involvement of private vendors in algorithmic design leads to another set of dangers, including opacity, public disempowerment, and loss of accountability. Public officials who have ceded the development of predictive algorithms to private vendors may not participate in and may be unaware of the policy decisions that are incorporated into those algorithms. Public employees who use the output of a predictive algorithm to inform their decisions may not understand the design and limitations

---

39 *See, e.g.*, Erika M. Kitzmiller, *IDS Case Study: Allegheny County*, ACTIONABLE INTELLIGENCE FOR SOC. POL'Y (May 2014) (analyzing how Allegheny County, Pennsylvania has used data analytics to improve its human service agency's responsiveness).

40 *See* Daniel Castro, *Data Detractors Are Wrong: The Rise of Algorithms is a Cause for Hope and Optimism*, CTR. FOR DATA INNOVATION (2016), http://www.datainnovation.org/2016/10/ data-detractors-are-wrong-the-rise-of-algorithms-is-a-cause-for-hope-and-optimism/ [http://perma.cc/YBS8-74P3].

41 On the Arnold Foundation's PSA-Court algorithms, concerning which we filed open records requests, see text accompanying notes 123-136.

of the algorithm, and may not be able to determine whether it has taken into account factors that they would consider relevant. Even if they were capable of interrogating the algorithm, they may be halted by vendor-drafted contracts that give the vendor control or ownership of the data and analytics.[42]

Private participation in public administration through algorithmic governance raises concerns that data will be used to hurt citizens and weaken public authority. "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."[43] 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."[44] The risk is that the corporation controlling the data and analytics occupies the command center of urban governance while the democratically accountable officials, unable to control the data, move to the periphery.[45]

A related concern is that, through these partnerships, private vendors come to lock governments into proprietary systems. Some smart city commentators warn that "smart" projects are simply vehicles to sell municipalities comprehensive data management systems owned and managed by the vendor.[46]

---

42 *See, e.g.*, Memorandum of Understanding Between the Laura and John Arnold Foundation and the Superior Court of California, County of San Francisco 3 (Aug. 4, 2015), http://www.muckrock.com/foi/san-francisco-city-and-county-3061/san-francisco-public-safety-assessment-court-30096/#file-113830 [http://perma.cc/73QC-QLX9]; *see also* Angwin et al., *supra* note 7 (describing the COMPAS contract).

43 Edwards, *supra* note 16, at 33.

44 David Meyer, *How One European Smart City is Giving Back Power to its Citizens*, ALPHR (July 10, 2017), http://www.alphr.com/technology/1006261/how-one-european-smart-city-is-giving-power-back-to-its-citizens [http://perma.cc/FM7D-YVX9] (quoting Francesca Bria, digital chief of Barcelona). Another problem is that "the business model is creating dependence on very few providers." *Id.* 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.*

45 Christine Richter & Linnet Taylor, *Big Data and Urban Governance*, *in* BIG DATA AND URBAN GOVERNANCE 175, 180 (J. Gupta et al. eds., 2015) ("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, *Thinking Critically*, *supra* note 6.

46 *See, e.g.*, ADAM GREENFIELD, AGAINST THE SMART CITY (2013); Donald McNeill, *Global Firms and Smart Technologies: IBM and the Reduction of Cities*, 40 TRANSACTIONS INST. BRIT. GEOGRAPHERS 562 (2015); Wiig, *The Empty Rhetoric*

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 cannot control. Lock-in may extend beyond technical systems to the physical infrastructure in which they are embedded. For example, the Alphabet subsidiary Sidewalk Labs is partnering with Waterfront Toronto to build from scratch "holistically" a mini-city of 800 acres as "a place that is enhanced by digital technology and data."[47] It is probable that Sidewalk Labs will gather data and use city data to make algorithmic predictions about desirable waterfront activity.[48] It is unclear what ownership or access the public will have to the data or related analytics, how the physical infrastructure will be governed, or whether public entities will be able to take control of the informational and physical assets should they wish to end the relationship with the private company.

II. DEFINING MEANINGFUL TRANSPARENCY: WHAT THE PUBLIC NEEDS TO KNOW

As artificial intelligence and algorithmic prediction come quickly to penetrate local governance, it would be desirable for the public to know what policy judgments the algorithms reflect and how well they perform in achieving the objectives set for them. This Part identifies the components of meaningful transparency in light of how an algorithm operates. Part V operationalizes the idea of meaningful transparency through specific disclosure practices that we recommend for public predictive algorithms.

A. What Are the Algorithm's Politics?

---

of the Smart City, *supra* note 33, at 535 ("[T]he 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.").

47 Darrell Etherington, *Alphabet's Sidewalk Labs To Turn Toronto into a Model Smart City*, TechCrunch (Oct. 17, 2017), http://techcrunch.com/2017/10/17/alphabets-sidewalk-labs-to-turn-toronto-area-into-a-model-smart-city/ [http://perma.cc/2BTE-ZSFY].

48 Laura Bliss, *When a Tech Giant Plays Waterfront Developer*, CITYLAB (Jan. 9, 2018), http://www.citylab.com/design/2018/01/when-a-tech-giant-plays-waterfront-developer/549590/ [http://perma.cc/6LKW-FTE9] (proposing a "digital layer" that will tie together and manage physical infrastructure and interactions between the infrastructure, city services, and people); *cf.* SIDEWALK LABS, *Project Vision* (Oct. 17, 2017), http://sidewalktoronto.ca/wp-content/uploads/2017/10/Sidewalk-Labs-Vision-Sections-of-RFP-Submission.pdf.

Algorithmic governance has a politics. Judgments are encoded in the algorithmic process at all stages.[49] These are judgments that at some level the public should know and speak to. However, algorithms positioned merely as the means to scientific truth can conceal the values embedded in the underlying models.[50] Moreover, when private vendors control algorithmic governance, the politics of algorithms recede behind private hedges. The idea that algorithms are a science without politics can obscure the stakes of their private control that are clearer in other areas of privatization, such as schools and prisons.

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."[51] There will be tradeoffs in implementing any policy goal. For example, even when implementing a goal as uncontroversial as reducing traffic wait time, a policymaker has to consider what risk to pedestrian safety is permissible in the service of traffic flow, and how to factor in the reduction of tailpipe emissions. 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.[52] We address specific contextualized problems in Part III. For now, it is enough to

---

49 Mittelstadt et al., *The Ethics of Algorithms: Mapping the Debate*, July-Dec. 2016 BIG DATA & SOC'Y 1, 1 ("Operational parameters are specified by developers and configured by users with desired outcomes in mind that privilege some values and interests over others.").

50 *See* Rob Kitchin, *Reframing, Reimagining and Remaking Smart Cities: The Programmable City* 4 (Open Sci. Framework, Working Paper No. 20, 2016) (summarizing smart city critiques).

51 Harry Surden, *Values Embedded in Legal Artificial Intelligence* 5 (Univ. Colo. Law Legal Studies, Research Paper No 17-17, Oct. 18, 2017), http://ssrn.com/abstract=2932333 [http://perma.cc/3RUC-54PP].

52 *See* Diakopoulos, *supra* note 6, at 400 (discussing the value choices embedded in data prioritization, classification, association, and filtering); *see also* EXEC. OFFICE OF THE PRESIDENT NAT'L SCI. & TECH. COUNCIL COMM. ON TECH., PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE (2016), http://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf [http://perma.cc/TZ2W-PJT7]; Bryce Goodman & Seth Flaxman, *EU Regulations on Algorithmic Decision-Making and a "Right to Explanation,"* ARXIV (2016), http://arxiv.org/pdf/1606.08813.pdf [http://perma.cc/4GDR-XQ5P]; Felicitas Kraemer et al., *Is There an Ethics of Algorithms?*, 13 ETHICS & INFO. TECH. 251 (2011).

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.[53] Nick Diakopoulos observes that algorithmic tuning "can privilege different 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."[54]

One of the few algorithms for which this tuning was revealed is 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.[55] Criminologist and statistician Richard Berk, who consulted on the program, estimates that between 29% and 38% of predictions end up being wrong—an error rate justified by a policy that it is "much more dangerous to release Darth Vader than it is to incarcerate Luke Skywalker."[56] 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

---

53 *See* Daniel Neyland & Norma Mollers, *Algorithmic IF . . . THEN Rules and the Conditions and Consequences of Power*, 20 INFO. COMM. & SOC'Y 45 (2016) (discussing algorithms that make just this kind of classification).

54 Diakopolous, *supra* note 6, at 401; *see also* Matthias Spielkamp, *Inspecting Algorithms for Bias*, MIT TECH. REV. (June 12, 2017), http://www.technologyreview.com/s/607955/ inspecting-algorithms-for-bias/ [http://perma.cc/K243-2DH2] (arguing that 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)).

55 Nancy Ritter, *Predicting Recidivism Risk: New Tool in Philadelphia Shows Great Promise*, 271 NAT'L INST. JUST. (2013), http://www.nij.gov/journals/271/pages/predicting-recidivism.aspx [http://perma.cc/J9SA-88LV].

56 Joshua Brustein, *This Guy Trains Computers to Find Future Criminals*, BLOOMBERG TECH. (July 18, 2016), http://www.bloomberg.com/features/2016-richard-berk-future-crime/ [http://perma.cc/46L5-4DMU]. *See generally* RICHARD A. BERK, STATISTICAL LEARNING FROM A REGRESSION PERSPECTIVE 13, 139-45 (2008) (discussing scoring of different kinds of errors in machine learning algorithms).

project participant, the model was intentionally made less accurate "to make sure it produces the right kind of error when it does."[57]

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."[58]

B. 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 the 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 an outcome, its predictions that subjects are among the riskiest ten percent are significantly less accurate than its predictions that subjects are among the least risky ten

---

57 *Id.* (quoting Geoffrey Barnes).

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

percent.[59] In other words, there are many ways to measure accuracy, with more coming down the pike.[60] One cannot understand claims about performance without knowing how it is being measured.

There are many reasons an algorithm could be ineffective.[61] It could be trained on bad data inputs (garbage in, garbage out).[62] Errors may also result from faulty inductive reasoning, data selection, and factor weighting.[63] Another point of failure in the broader algorithmic process may be at the implementation phase. Unless the algorithmic prediction is self-executing, human beings have to understand the prediction in order to choose how much weight to give it in the decision-making process. 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 to enable proper evaluation.

C. Is the Algorithm Fair?

---

59 *See, e.g.*, Alberto Maydeu-Olivares & C. Garcia-Forero, *Goodness-of-Fit Testing*, *in* INTERNATIONAL ENCYCLOPEDIA OF EDUCATION 190 (Penelope Peterson et al. eds., 3d ed. 2010).

60 *See, e.g.*, DEAN ABBOTT, APPLIED PREDICTIVE ANALYTICS: PRINCIPLES AND TECHNIQUES FOR THE PROFESSIONAL DATA ANALYST 283-304 (2014); Ewout W. Steyerberg et al., *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), http://doi.org/10.1186/1471-2164-13-S4-S2 [http://perma.cc/8Q73-GXYC]; Scott Fortmann-Roe, *Accurately Measuring Model Prediction Error* (May 2012), http://scott.fortmann-roe.com/docs/MeasuringError.html [http://perma.cc/3YDK-QDMK].

61 *See* EXEC. OFFICE OF THE PRESIDENT, *httphttpsupra* note 24 (discussing poorly selected data; incomplete, incorrect or outdated data; selection bias; and unintentional perpetuation and promotion of historical biases).

62 *Garbage In, Garbage Out*, WIKIPEDIA, http://en.wikipedia.org/wiki/Garbage_in,_ garbage_out [http://perma.cc/M8CV-P7U8] (explaining 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).

63 *See* Amended Summary Judgment Op., Hous. Fed'n of Teachers v. Hous. Indep. Sch. Dist., Civil Action No. H-14-1189, at *13 (S.D. Tex. 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.").

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.[64]

Government use of predictive algorithms poses an inherent challenge to traditional notions of fairness.[65] By their nature, predictive models are simplifications that cannot consider all possible relevant facts about subjects, and that therefore necessarily treat people as members of groups, not as individuals.[66] Generalizations, which may treat unlike cases alike, are inherent to this process. For sensitive decisions, particularly where individual liberty is at stake, decision makers like judges and social workers are expected to exercise human judgment over algorithmic predictions so that they may catch faulty predictions. In theory, the algorithmic edict is advisory only. In practice, decision makers place heavy reliance on the numbers, raising the stakes for their fairness.[67]

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.[68] Additional forms of discrimination are of concern, such as whether an algorithm incidentally disfavors (and therefore, disincentivizes) certain behaviors. For example, if use of the mental health system

---

64 Nick Seaver, *Knowing Algorithms* 2 (Feb. 2014) (unpublished manuscript), http://nickseaver.net/s/seaverMiT8.pdf [http://perma.cc/AB75-JY6M] (arguing that 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*, *in* PROC. 21ST ACM SIGKDD INT'L CONF. DISCOVERY & DATA MINING 259 (2015) (presenting a test for disparate impact in algorithmic processes and a method by which data might be made unbiased); Sorelle A. Friedler et al., *On the (Im)possibility of Fairness* (Sept. 23, 2016) (unpublished manuscript), http://arxiv.org/pdf/1609.07236.pdf [http://perma.cc/DQP4-PMWU].

65 Fairness itself is subject to different definitions; the definition selected will affect assessments of algorithmic fairness. *See* Friedler et al., *supra* note 64 (recommending that computer scientists make more explicit what notion of fairness they seek to represent in algorithms).

66 *See, e.g.*, O'NEIL, *supra* note 6, at 20-23; Mittelstadt et al., *supra* note 49, at 8.

67 *See generally* John Danaher, *The Threat of Algocracy: Reality, Resistance and Accommodation*, 29 PHIL. & TECH. 245 (2016) (raising concerns about deference to algorithmic output by human decision makers who cannot understand how the algorithms work).

68 *See, e.g.*, O'NEIL, *supra* note 6; Barocas & Selbst, *supra* note 7; Joh, *supra* note 4; Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857 (2017); Andrew D. Selbst, *Disparate Impact in Big Data Policing*, GA. L. REV. (forthcoming), http://ssrn.com/abstract=2819182 [http://perma.cc/KDT8-DXF5].

correlates with increased risk of child endangerment, then an algorithm 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 are sometimes correlated.[69] A classic example is facial recognition algorithms that were trained on the faces familiar to the engineers who built it, which were mostly white.[70] As a result, the programs were more likely to fail to identify or to misidentify dark-skinned human faces, which may make innocent dark-skinned people more likely to be misidentified as criminal suspects.[71] Retraining the algorithm using human faces of all skin colors would make it perform better overall, as well as reduce the disparity in accuracy 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 will 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.[72] That correlation might not increase the

---

69 *See, e.g.*, Tal Z. Zarsky, *Understanding Discrimination in the Scored Society*, 89 WASH. L. REV. 1375, 1383 (2014) ("[I]n many instances, discrimination might be inefficient and thus present an unsustainable business or social practice").

70 *See* Tess Townsend, *Most Engineers Are White—And So Are the Faces They Use To Train Software*, RECODE (Jan. 18, 2017), http://www.recode.net/2017/1/18/14304964/data-facial-recognition-trouble-recognizing-black-white-faces-diversity [http://perma.cc/T5PL-XNLJ]. *See generally* Clare Garvie & Jonathan Frankle, *Facial-Recognition Software Might Have a Racial Bias Problem*, ATLANTIC (Apr. 7 2016), http://www.theatlantic.com/technology/ archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991/ [http://perma.cc/6K5Z-H3JM] (describing the racial biases of facial recognition algorithms).

71 *See* Garvie & Frankle, *supra* note 70.

72 *See* Angwin et al., *supra* note 7. *But see* William Dieterich et al., *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*, NORTHPOINTE INC. RES. DEP'T (July 8, 2016), http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_ 070616.pdf [http://perma.cc/ND8Z-Y3GS]; Anthony W. Flores et al., *False 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*

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 co-variants, but with employment history as the better predictor. It might be the case 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.

However, due to residential segregation, zip codes are often proxies for race. Knowing this, agencies may choose to exclude zip codes as inputs to predictive algorithms even where they improve the algorithm's predictive power. They may conclude that skin color has no causal relation to the desired or undesired outcome and, therefore, that the predictive power must be rooted in some other co-variant. To use race, or its proxy, as a shortcut for whatever might actually have some causal relation would perpetuate "a history of purposeful unequal treatment"[73] based on "an immutable characteristic determined solely by the accident of birth."[74] 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.[75] 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

---

Blacks," COMMUNITY RESOURCES FOR JUST., http://www.crj.org/page/-/publications/rejoinder7.11.pdf [http://perma.cc/624E-ZUE9]; Rhema Vaithianathan et al., *Developing Predictive Models to Support Child Maltreatment Hotline Screening Decisions: Allegheny County Methodology and Implementation*, CTR. FOR SOC. DYNAMICS (Sept. 2016) (discussing zip codes and other proxies for race).

73 San Antonio Ind. Sch. Dist. v. Rodriguez, 411 U.S. 1, 28 (1973).

74 Frontiero v. Richardson, 411 U.S. 677, 686 (1973).

75 *See* O'NEIL, *supra* note 6, at 97-98.

recorded."[76] 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, particularly data concerning drug-related and nuisance crimes, in creating its policing models, to avoid likely systemic bias.[77] Ultimately, unless a predictive algorithm is rendered sufficiently transparent, we will not know whether automated decision-making and risk prediction accords with our substantive commitments to fairness.[78]

D. Does the Algorithm Enhance or Diminish Governmental Capacity?

There is a further danger that algorithmic governance, impervious to critical evaluation while also displacing human decision making, will hollow out the decision-making capacity of public servants. Contributing factors could include unwarranted deference to the algorithm, insufficient understanding of algorithmic processes, and 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 decision makers to resist the edict. As Harry Surden notes, judges may "give more deference to computer-based recommendations, in contrast to comparable human-based assessments, given the aura of mechanistic objectivity surrounding computer-generated, analytics-based, analyses."[79] According to Michael Ananny, "algorithmic categories . . . signal certainty, discourage alternative explorations, and create coherence among disparate objects,"

---

76 Emily Thomas, *Why Oakland Police Turned Down Predictive Policing*, VICE (Dec. 28, 2016), http://motherboard.vice.com/read/minority-retort-why-oakland-police-turned-down-predictive-policing [http://perma.cc/KJH5-DS93].

77 *See A Citizen's Guide to HunchLab*, HUNCHLAB 26 (July 11, 2017), http://robertbrauneis.net/algorithms/HunchLabACitizensGuide.pdf [http://perma.cc/NY6B-QNZD]. 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.

78 See Joanna Bryson, *Three Very Different Sources of Bias in AI, and How To Fix Them*, ADVENTURES IN NI (July 13, 2017), http://joanna-bryson.blogspot.co.uk/2017/07/three-very-different-sources-of-bias-in.html [http://perma.cc/2RXD-LZPG] ("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.").

79 *See* Surden, *supra* note 51, at 2; *see also* danah boyd & Kate Crawford, *Critical Questions for Big Data*, 15 INFO. COMM. SOC'Y 662 (2012) (identifying mistaken belief in objectivity as one of the pitfalls of reliance on big data analytics).

thereby reifying the algorithmic model's choices.[80] 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 are 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 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."[81]

Over time, deference to algorithms may weaken the decision-making 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,[82] and it will undoubtedly spread to public administration. Ethicists have also examined how computer systems can undermine a person's sense of her 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."[83] 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 algorithms 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.[84]

---

80 Ananny, *supra* note 6 at 103; *see also* Barocas & Selbst, *supra* note 7.

81 Christopher E. Church & Amanda J. Fairchild, *In Search of a Silver Bullet: Child Welfare's Embrace of Predictive Analytics*, 68 JUV. & FAM. CT. J. 67, 78 (2017).

82 *See, e.g.*, NICHOLAS CARR, THE GLASS CAGE: HOW OUR COMPUTERS ARE CHANGING US 106-112 (2014).

83 Batya Friedman & Peter H. Kahn, Jr., *Human Agency and Responsible Computing: Implications for Computer System Design*, 17 J. SYSTEMS SOFTWARE 9 (1992); *see* Helen Nissenbaum, *Accountability in a Computerized Society*, 2 SCI. & ENGINEERING ETHICS 25 (1996).

84 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 decision makers, see Gwen Robinson, *Implementing OASys:*

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 decision making 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.

E. Meaningful 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 works. Algorithmic opacity is a problem widely recognized and variously defined.[85] As a general matter[86] and with respect to public sector applications, commentators recognize the need for more transparency in the implementation of predictive algorithms.[87] So do courts presented with cases of first

---

Lessons from Research into LSI-R and ACE, 50 PROBATION J. 30, 33 (2003); and Diana Wendy M. Fitzgibbon, *Risk Analysis and the New Practitioner: Myth or Reality?*, 9 PUNISHMENT & SOC'Y 87, 90 (2007)

85 For an exploration and taxonomy of various kinds of algorithmic opacity, see Andrew D. Selbst & Salon Barocas, Regulating Inscrutable Systems (Mar. 20, 2017) (unpublished manuscript), http://www.werobot2017.com/wp-content/uploads/2017/03/Selbst-and-Barocas-Regulating-Inscrutable-Systems-1.pdf [http://perma.cc/Y5FZ-E6SU].

86 *See, e.g.*, Frank Pasquale, *Restoring Transparency to Automated Authority*, 9 J. TELECOMM. & HIGH TECH. L. 235 (2011)**;** Julie Brill, Comm'r, Fed. Trade Comm., Scalable Approaches to Transparency and Accountability in Decisionmaking Algorithms: Remarks at the NYU Conference on Algorithms and Accountability (Feb. 28, 2015), http://www.ftc.gov/system/files/documents/public_statements/629681/150228nyualgorithms.pdf [http://perma.cc/39XN-YN42]; Nicholas Diakopoulos, *Revealing Algorithms*, ELECTRONIC PRIVACY INFO. CTR., http://epic.org/algorithmic-transparency/ [http://perma.cc/L3X8-XJ8C]; Ed Felten, *Accountable Algorithms*, FREEDOM TO TINKER (Sept. 12, 2012), http://freedom-to-tinker.com/2012/09/12/accountable-algorithms/ [http://perma.cc/3SM8-N2TN];httphttphttphttp. *But see* Mike Ananny & Kate Crawford, *Seeing Without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability*, NEW MEDIA & SOC'Y (Dec. 13, 2016), http://journals.sagepub.com/doi/abs/10.1177/1461444816676645 [http://perma.cc/9TS6-837G] (arguing that some algorithmic processes may be inherently non-transparent and, therefore, should be rendered accountable in other ways).

87 Lee P. Breckenridge, *Water Management for Smart Cities: Implications of Advances in Real-Time Sensing, Information Processing, and Algorithmic Controls*, 7 GEO.WASH. J. ENERGY & ENVTL. 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 . . .

impression about the due process rights of individuals affected by algorithmic judgment to know the reasons why the machine "said so."[88]

To be sure, there has always been risk of inefficacious or biased decision making 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 see the decision-making processes that are unknowable in a human subconscious and to test whether those predictions are inaccurate or unfair. 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—and even across many local and state governments.[89] This is the problem that Cathy O'Neil identifies as the scalability of algorithms.[90] 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. Part V below identifies 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

---

unless the administrative processes for adopting these systems are themselves made accessible, transparent, and subject to ongoing and meaningful review").

88 *See, e.g.*, Amended Summary Judgment Op., Hous. Fed'n of Teachers v. Hous. Indep. Sch. Dist., Civil Action No. H-14-1189 (S.D. Tex. May 4, 2017) (allowing teachers' due process action against public school district for implementing a teacher evaluation algorithm that is impervious to investigation).

89 For example, the Arnold Foundation's Public Safety Assessment is used by three entire states—Arizona, Kentucky, and New Jersey—and in approximately thirty-five other jurisdictions. *See Public Safety Assessment*, LAURA & JOHN ARNOLD FOUND., http://www.arnoldfoundation.org/initiative/criminal-justice/crime-prevention/public-safety-assessment/ [http://perma.cc/9XLV-N65H].

90 *See* O'NEIL, *supra* note 6, at 29-31.

formulas, inputs, and rules of procedure in a "black box."[91] 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 formal disclosure in some combination of mathematical and logical notation and natural language.[92] 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.[93] 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 to easier means of interpretation.

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?[94] Access to the underlying data or at least descriptions of it would help us understand how strong the purported correlations actually are,

---

91 *See* PASQUALE, *supra* note 6, at 1-18; *see also* O'NEIL, *supra* note 6, at 28-31 (illustrating the ways in which creators of predictive algorithms conceal their mechanics); Selbst & Barocas, *supra* note 85, at 9 (outlining the "What-How-Why" model of explanation predictive algorithms fail to satisfy).

92 For one example of such disclosure, see MARIE VANNOSTRAND & GINA KEEBLER, PRETRIAL RISK ASSESSMENT IN THE FEDERAL COURT 48 (Apr. 14, 2009), http://tiny.cc/r3qrmy [http://perma.cc/YKZ5-6FV6] (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).

93 *See* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1268-69 (2008) (stating that from September 2004 to April 2007, code writers embedded over nine hundred 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).

94 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. Amended Summary Judgment Op., Hous. Fed'n of Teachers v. Hous. Indep. Sch. Dist., Civil Action No. H-14-1189, at *17 (S.D. Tex. May 4, 2017).

what the sample size was, and other matters that affect statistical validity.

Other types of information now typically sunk in obscurity include 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."[95] If an algorithm uses hundreds of unweighted inputs in a complex decision tree in which a single input might appear at multiple junctures, we cannot necessarily figure out which inputs were decisive in a particular case.[96] 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 to ignore their own judgment, and perhaps eventually to lose that judgment.

Lastly, algorithmic processes may be dynamic, their rules changing constantly to fit new patterns in the data. As a result, the code and data sets that are released to the public at Time 1—even if "interpretable" in isolation—may bear little resemblance to the process that is conducted at Time 2. Dynamic algorithms are, as Rob Kitchin says, "ontogenetic in nature," subject to being "edited, revised, deleted and restarted."[97] This dynamism creates obstacles to transparency, as the algorithms

---

95 Of course, people can sometimes provide us with explanations of their reasoning processes. 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* (June 16, 2016) (unpublished manuscript) (manuscript at 98), http://arxiv.org/abs/1606.03490v2 [http://perma.cc/H3E5-TVA2] (noting this, and defining the ability of a model to be explained after the fact as "post hoc interpretability").

96 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 is doing. Suppose, however, that the model uses over one thousand factors to predict parolee recidivism, some of which do not 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).

97 Kitchin, *Thinking Critically*, *supra* note 6, at 18.

themselves become less easy to understand.[98] 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.[99] It is merely a means. 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.[100] In the public sphere, this entails that government actually be held accountable by the voting public for the algorithms it deploys. Such accountability requires not *perfect* transparency—complete knowledge of an algorithm's rules of operation and process of creation and validation—but the lower standard of *meaningful* transparency—knowledge sufficient to approve or disapprove of the algorithm's performance. Records short of the underlying computer code may suffice to provide the necessary input. Of course, accountability in practice 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.

III. AN OPEN RECORDS ACT PROJECT: OBTAINING DOCUMENTATION OF ALGORITHMS

Open data practices are probably the best way to make transparent aspects of the algorithms used in government.[101] 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 from the government, which reduces the load on "pull" requests for government records from the public.[102] Yet governments have not been pushing out information about the algorithms they use. There is a big gap between the importance of algorithmic

---

98 *See* EXEC. OFFICE OF THE PRESIDENT, *supra* note 24 **(**noting that 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").

99 Kroll et al., *supra* note 26, at 657-60; *see also* Ananny, *supra* note 6, at 109.

100 *See* Kroll et al., *supra* note 26, at 657-60; Selbst & Barocas, *supra* note 85, at 15.

101 *See* JOSHUA TAUBERER, THE PRINCIPLES AND PRACTICES OF OPEN GOVERNMENT DATA 10 (2d ed. 2014).

102 The FOIA Improvement Act of 2016 amended 44 U.S.C. § 3102 (requiring now that 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").

processes for governance and public access to those algorithms.[103] 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. We first discuss our project design, introducing open records laws and explaining how we chose which algorithms and public agencies to target and how we formulated our open records requests. We then discuss our results.

A. Project Design and Implementation

1. 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.[104]

Freedom of information laws have as their principal goal accountable government. In signing the original federal 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."[105] 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."[106] And when Congress affirmed in 1996 that the central transparency mandate of FOIA applied to electronic records, the Senate Committee Report explained that

---

103 *See* Nicholas Diakopoulos, *Algorithmic Accountability Reporting: On the Investigation of Black Boxes*, TOW CTR. FOR DIGITAL JOURNALISM 2 (2014), http://academiccommons.columbia.edu/catalog/ac:2ngf1vhhn4 [http://perma.cc/7N2Q-7XHA] ("What we generally lack as a public is clarity about how algorithms exercise their power over us.").

104 *See, e.g.*, New York Freedom of Information Law, N.Y. PUB. OFF. L. § 86(4) (McKinney 2003) ("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.").

105 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), http://www.justice.gov/oip/67agmemo.htm [http://perma.cc/XD4G-FLNR].

106 S. REP. NO. 93-854, at 1 (1974).

government transparency "is consistent with our democratic form of government by furthering the interests of citizens in knowing what their Government is doing."[107] The courts have consistently held that FOIA embodies "a general philosophy of full agency disclosure."[108]

Animated by the same transparency principles, the open records statutes of all fifty 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.[109] Almost all the laws also exempt trade secrets, which we discuss below.

FOIA only applies to "agency records"—a term undefined in the statute.[110] 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.[111] Although state laws more typically use the term "government records," the coverage is similar.[112]

FOIA covers digital records, including software and databases.[113] Some state laws expressly include software as a public record.[114] 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

---

107 S. REP. NO. 104-272, at 5 (1996). *See generally* MICHAEL SCHUDSON, RISE OF THE RIGHT TO KNOW (2015) (chronicling the history of FOIA).

108 Dep't of the Air Force v. Rose, 425 U.S. 352, 360 (1976).

109 *See, e.g.*, 5 U.S.C. §§ 552(b)(1)-(9) (2012). 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).

110 5 U.S.C. § 552(a)(4)(B).

111 *See* Forsham v. Harris, 445 U.S. 169, 182 (1980).

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

113 *See, e.g.*, 5 U.S.C.§(f)(2)(A) (providing that "'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.*, CAL GOV'T CODE §§ 6254.9(a)-(b) (excluding computer software "developed by a state or local agency . . . includ[ing] computer mapping systems, computer programs, and computer graphic systems")

114 *See generally* Cristina Abello, Reporters Committee for Freedom of the Press, *Access to Electronic Communications*, REPORTERS COMMITTEE FOR FREEDOM PRESS (2009), http://www.rcfp.org/rcfp/orders/docs/ELECCOMM.pdf [http://perma.cc/QBM4-UCZ8]; 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).

official business."[115] Other state statutes expressly *exclude* software from public records.[116] Still others have not addressed the issue.[117] 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 to disclosure is 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.[118] 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.[119] 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.

2. Algorithms, Agencies, and the Formulation of Requests

---

115 N.J. STAT. ANN. § 47:1A-1.1 (West 2015); *cf.* FLA. STAT. § 119.011(12) (2017) (defining "public records" to include "data processing software"); 5 ILL. COMP. STAT. 140/2(c) (2016) (defining "records" as including "electronic data processing records."); N.Y. PUB. OFF. L. § 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").

116 *See, e.g.*, ALASKA STAT. § 40.25.220(3) (2010) (providing that the definition of "public records" does not include "proprietary software programs")

117 *See, e.g.,* ALA. CODE § 36-12-40 (2004) (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 the meaning of "writing").

118 *See* Alexandra Mateescu, et al., Police Body-Worn Cameras 9 (Data & Soc'y Research Institute, Working Paper, Feb. 2, 2015), http://www.datasociety.net/pubs/dcr/PoliceBodyWornCameras.pdf [http://perma.cc/TB9A-K63G] (discussing police department storage of police body camera footage in third-party cloud servers).

119 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 L.J. 19, 27 (2013).

Although our open records project was "empirical" in the sense that we sent requests out into the world to see how governments would respond, it was not and could not have been quantitative or statistical. There is no central registry of algorithms in use by governments, and algorithms are not naturally visible in the way that, say, skyscrapers or bridges are. Thus, we have no means of knowing how many algorithms are currently in use, who has developed them, or which governments are using them. Without that knowledge, we cannot develop any method for sampling algorithm use in any way that would allow us to generalize from our findings.

What we have done is much less formal. We surveyed public information to identify local government use of predictive algorithms. We then chose six programs to get a mix of different subject matters (policing, criminal justice, child welfare, and education), of different developers (foundations, private corporations, and government entities), and of different jurisdictions (forty-two different agencies in twenty-three states). The six programs are: Public Safety Assessment; Eckerd Rapid Safety Feedback; Allegheny Family Screening Tool; PredPol; HunchLab; and New York City Value-Added Measures. We drafted an open records request that was by design very general and inclusive, trying to cover any record that related to the algorithm in question.[120] When an agency responded that the breadth of the request was causing delays, we sent a revised request.[121] Because of time and resource constraints, we did not challenge final denials or sustained non-responsiveness in court, nor did we pay significant fees when an agency demanded them to respond to our request.[122] The barriers we encountered amount to substantial limitations on public access to information about algorithms, even if some of them could be overcome with more time and money. In some cases, we were also able to engage in direct communication with algorithm developers to try to gain additional information.

---

120 For one example of our standard initial request language, see our request to the Lincoln, Nebraska Police Department regarding its use of HunchLab, Letter from Michael Morisy to Lincoln Police Dep't (Nov. 15, 2016), http://www.muckrock.com/foi/lincoln-4033/lincoln-police-department-hunchlab-documents-30110/ [http://perma.cc/Z8LY-RT32].

121 See, for example, our correspondence with the Cocoa (Florida) Police Department, Letter from Michael Morisy to Cocoa Police Dep't (Nov. 16, 2016), http://www.muckrock.com/foi/cocoa-10491/cocoa-police-department-predpol-documents-30104/ [http://perma.cc/3KEA-HJF3].

122 *See, e.g.*, Letter from Okla. Dep't of Human Servs. to Muck Rock [sic] (Nov. 3, 2016), http://www.muckrock.com/foi/oklahoma-248/oklahoma-department-of-human-services-eckerd-rapid-safety-feedback-28151/#file-108045 [http://perma.cc/Q88A-8JVQ] (requesting payment of $2,472 before work would begin on responding to our request).

B. Results

In general, we found wide variation in whether jurisdictions responded to our requests; 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 were developed. Some developers were also more forthcoming than others. The Arnold Foundation, developer of Public Safety Assessment, has disclosed its relatively simple algorithms to the public, but 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. We proceed to detail the responses with regard to each of the six algorithms.

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.[123] As of this writing, it is being used in thirty-eight jurisdictions, including the entire states of Arizona, New Jersey, and Kentucky.[124] 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; or 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 signifies lowest risk and six signifies 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.[125]

---

123 *Public Safety Assessment*, *supra* note 89.

124 *See id.*

125 A description of all three algorithms, including factors, raw point allocations, and conversion from raw scores to final output, is available at *Public Safety Assessment: Risk Factors and Formula*, LAURA & JOHN ARNOLD FOUND. (2013-

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?

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 sixteen 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,[126] 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.[127] Three of

16), http://www.arnoldfoundation.org/wp-content/uploads/PSA-Risk-Factors-and-Formula.pdf [http://perma.cc/GNF5-89PV].

126 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 from Michelle M. Smith, Clerk of the Superior Court, to Michael Morisy (Dec. 22, 2016), http://www.muckrock.com/foi/new-jersey-229/new-jersey-superior-court-public-safety-assessment-court-28835/#file-114392 [http://perma.cc/M9ZZ-JU3G]. 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. E-mail from Christopher H. Connors, Chief Deputy Court Adm'r, to Michael Morisy (Oct. 6, 2016), http://www.muckrock.com/foi/allegheny-county-306/allegheny-county-public-safety-assessment-court-28150/ [http://perma.cc/3BYS-4RDM].

127 *See* Letter from Ann E. Donlan, Commc'ns Dir., Superior Court of Cal., Cty. of San Francisco, to Michael Morisy (Dec. 16, 2016), http://www.muckrock.com/foi/san-francisco-city-and-county-3061/san-francisco-public-safety-assessment-court-30096/#file-113829 [http://perma.cc/A48W-RN5A]; Letter from Ronald G. Overholt, Court Adm'r,

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.[128]

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.[129]

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%), and 6 (40%).[130] Thus, the highest risk score was set to generate

---

Ariz. Superior Court, Pima Cty., to Michael Morisy (Jan 18, 2017), http://www.muckrock.com/foi/pima-county-183/pima-county-superior-court-public-safety-assessment-30130/#file-116730 [http://perma.cc/EER2-PS5L]; E-mail from Marla Randall, Court Adm'r, Navajo Cty. Courts, to Michael Morisy (Dec. 20, 2016), http://www.muckrock.com/foi/navajo-county-9526/navajo-county-superior-court-public-safety-assessment-30129/ [http://perma.cc/YXB9-2454]; E-mail from Paul Thomas, Court Adm'r, Mesa Mun. Court, to Michael Morisy (Nov. 17, 2016), http://www.muckrock.com/foi/mesa-4736/mesa-municipal-court-public-safety-assessment-30126/ [http://perma.cc/V6P5-ZAP5].

128 *See, e.g.*, Memorandum of Understanding Between the Laura and John Arnold Foundation and the Superior Court of California, County of San Francisco, *supra* note 42 ("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.").

129 Memorandum of Understanding Between the Laura and John Arnold Foundation and the Seventh Judicial Circuit of the State of Florida 3 (2015), http://cdn.muckrock.com/foia_files/2016/12/15/Memorandum_of_Understanding.pdf [http://perma.cc/4QAE-NUCP].

130 *See* Zach Dal Pra, LJAF Public Safety Assessment—PSA, JUST. SYS. PARTNERS 31, http://cdn.muckrock.com/foia_files/2016/12/15/Volusia_Stakeholder_Training_10162015.pdf [http://perma.cc/BH7A-PNPC] (slide deck). 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.

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%), and 6 (30%).[131] 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.[132] 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.[133] 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 systems.[134] 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.

---

131 *Id.* at 46 (finding these numbers based on tracking PSA application in 100,000 cases in Kentucky and in three unnamed cities outside of Kentucky).

132 *See* Dal Pra, *supra* note 130, slides 49-54 (describing the Decision Making Framework); Zach Dal Pra, *Volusia County Case Review*, JUST. SYS. PARTNERS 3 (2016), http://cdn.muckrock.com/foia_files/2016/12/15/Volusia_County_Case_Review_Redacted.pdf [http://perma.cc/A4PA-2EQR] (noting that "[p]retrial staff is adhering to the Decision-Making Framework as a guide to making recommendations that are based on the defendant's risk levels with adjustments for the nature of the present charge").

133 *See* Dal Pra, *supra* note 130, slide 52 (explaining how the Decision Making Framework is used to "determine the preliminary recommendation release type and corresponding conditions level.").

134 *See* Email from Leila Walsh, Vice President of Commc'n, Laura & John Arnold Found., to Robert Brauneis, Professor of Law, George Washington Univ. 5 (Mar. 2, 2017), http://www.robertbrauneis.net/algorithms/ArnoldFoundationEMail.pdf [http://perma.cc/7CU3-4JQP] ("Before implementing the PSA, local stakeholders—such as representatives from the courts, law enforcement, district attorney's office, and public defender's office—work together on a decision-making framework (DMF) for their jurisdiction.").

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.[135]

From the Foundation's website, the documents provided by the Seventh Judicial Circuit, and the statement the Foundation produced for us, we know 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 memoranda of understanding 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.[136]

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.[137] RSF was developed by Eckerd Kids (Eckerd), a nonprofit family and child services

---

135 *Id.*

136 *Id.* at 5.

137 *Summary and Replication Information*, ECKERD RAPID SAFETY FEEDBACK, http://static.eckerd.org/wp-content/uploads/Eckerd-Rapid-Safety-Feedback-Final.pdf [http://perma.cc/6H3D-PNBR].

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 violence history, and a parent who had previously been placed in foster care."[138] Eckerd partnered with Mindshare Technology to create software that analyzes data in existing child welfare reporting systems and flags high-risk cases with these factors for intervention.[139]

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.[140] Maine acknowledged our request but to date has not produced any documents.[141]

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 Memorandum 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

---

138 *Id.*; *see also* Bryan Lindert, *Eckerd Rapid Safety Feedback: Bringing Business Intelligence to Child Welfare*, 2014 POL'Y & PRAC. 25, http://static.eckerd.org/wp-content/uploads/Eckerd.pdf [http://perma.cc/SKU8-UJQH]. 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.* It is not clear whether Eckerd performed these analyses by means of machine learning, human evaluation, or some combination of the two.

139 *See Summary and Replication Information*, *supra* note 137.

140 *See* Invoice, Open Record Request, OKLA. DEP'T HUMAN SERVICES (Nov. 3, 2016), http://cdn.muckrock.com/foia_files/2016/11/09/11-3-16_MR28151_FEE_2472.pdf [http://perma.cc/L78B-DRBK] (requesting payment of $2472 to respond to open records request).

141 *See* Letter from Kevin C. Wells, General Counsel, Me. Dep't of Health and Human Servs. to Michael Morisy, MUCKROCK (Nov. 21, 2016), http://cdn.muckrock.com/foia_files/2016/12/06/1249s5dayletterEckerdRapid_SafetySoftware11.21.16.pdf [http://perma.cc/34RS-TZLX].

maintained by Eckerd, to which OCS personnel can gain access.[142]

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 Memorandum of Understanding 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.[143]

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.[144]

The Illinois Department of Children and Family Services provided its contracts with Eckerd for Fiscal Years 2016 and 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.[145] The contracts also contain what appears to be standard state contracting provisions that are more favorable to disclosure and public ownership than the Alaska Memorandum

---

142 Memorandum of Understanding of February 20, 2015 between Eckerd Youth Alternatives, Inc. and the Alaska Department of Health and Social Services 2, http://cdn.muckrock.com/foia_files/2017/03/02/MOU_Eckerd_and_OCS_February_2015_signed.docx.pdf [http://perma.cc/58KW-FXKP] (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").

143 *See id.* at 1, 4, 5.

144 For example, a chart entitled "CT Eckerd Rapid Safety Feedback Process Flow" allocates to Mindshare the step of "Mindshare Tool/Predictive Analysis Generates List for Review." *CT Eckerd Rapid Safety Feedback Process Flow*, ECKERD, http://cdn.muckrock.com/foia_files/2016/11/13/CT_ERSF_Process_Flowchart9-12.pdf [http://perma.cc/4B3Z-ZVPG].

145 *See* Contract # 5445089016 between Eckerd Youth Alternatives, Inc. and State of Illinois, Department of Children and Family Services 7 (Sept. 15, 2015), http://www.robertbrauneis.net/algorithms/ILERSFFY16.pdf [http://perma.cc/D5LY-44C5] [hereinafter Illinois FY16 Contract]; Contract # 5445089027 between Eckerd Youth Alternatives, Inc. and State of Illinois, Department of Children and Family Services 7 (July 1, 2016) , http://www.robertbrauneis.net/algorithms/ILERSFFY17.pdf [http://perma.cc/M7QS-XLH7] [hereinafter Illinois FY17 Contract].

of Understanding. They recite that "[a]ny information not prohibited or exempt from disclosure under federal law, State law, or applicable FOIA exemption is public."[146] They also provide that the state owns everything produced under the contracts, including all intellectual property rights and any products of the contracts.[147] 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 "[t]he development of the predictive model that will be used to identify those incoming investigations with the highest probability of serious injury or death."[148] 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 a request for proposals 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,[149] and indicated that they were 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

---

146 *See* Illinois FY16 Contract, *supra* note 145, at 11; Illinois FY17 Contract, *supra* note 145, at 11.

147 *See* Illinois FY16 Contract, *supra* note 145, at 11; Illinois FY17 Contract, *supra* note 145, at 11.

148 Illinois FY16 Contract, *supra* note 145, at 6.

149 *See* Vaithianathan et al., *supra* note 72.

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.[150] 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 eighteen months before that allegation, and two years after the allegation.[151] The dataset included over 800 variables.[152] 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.[153] 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).[154] 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.[155] 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.[156] 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.[157]

---

150 *See id.* at 10.
151 *See id.* at 11.
152 *See id.* at 12.
153 *See id.* at 13-14.
154 *See id.* at 15-17, 19-23.
155 *See id.* at 37-43. We made such a request and were provided with the weights.
156 *See id.* at 27.
157 *See id.* at 28.

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.[158] 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.[159]

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 the University of California, Los Angeles and Santa Clara University, in collaboration with crime analysts and officers from the Los Angeles and Santa Cruz Police Departments,[160] but is now managed by the 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.[161] PredPol feeds data about past patterns of criminal activity into an algorithm that predicts where and when new crimes will be committed.[162] According to one source, PredPol "is well known for keeping its algorithm a 'closely guarded' secret."[163]

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, acknowledged our request but did not

---

158 *See id.* at 15, 30.

159 *See id.* at 10.

160 *PredPol Is Predictive Policing*, PREDPOL, http://www.predpol.com/about/ [http://perma.cc/Y2R8-C2XM].

161 *Id.*

162 *Id.*

163 Elizabeth E. Joh, *The Undue Influence of Surveillance Technology Companies on Policing*, 92 N.Y.U. L. REV. ONLINE 101, 119 (2017) (quoting Ali Winston, *Arizona Bill Would Fund Predictive Policing Technology*, REVEAL (Mar. 25, 2015), http://www.revealnews.org/article/arizona-bill-would-fund-predictivepolicing-technology/ [http://perma.cc/V4BV-YMZP]).

produce documents, 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 two-hundred email threads of correspondence between Tacoma Police Department and PredPol personnel concerning a wide variety of issues in implementing PredPol. It also produced ten 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, 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."[164] 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 twenty-four-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

---

[164] Legislation Details (With Text), File # 15-361, City of Cocoa 1 (July 30, 2015), http://cdn.muckrock.com/foia_files/2017/01/13/15-361_City_Council_Agenda_Item__8-25-15.pdf [http://perma.cc/K9JZ-ZAR9].

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.[165] HunchLab uses a wide range of inputs to predict risks of crime, and allows individual police departments to prioritize for selected crimes.[166]

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.

Miami eventually responded that it had no responsive documents.[167] 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.[168] The City of Philadelphia produced a purchase order for the HunchLab service,[169] but otherwise denied our request on the grounds that we did not request specific documents.[170] 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 thirty-day rolling

---

165 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?*, B CORPS, http://www.bcorporation.net/what-are-b-corps [http://perma.cc/Q5FH-8KCK].

166 "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." Maurice Chammah, *Policing the Future*, MARSHALL PROJECT (Feb. 3, 2016), http://www.themarshallproject.org/2016/02/03/policing-the-future#.vVL53xF4m [http://perma.cc/YKC9-CDU6] (reporting on St. Louis' deployment of the HunchLab system).

167 *See Miami PD HunchLab Documents*, MUCKROCK, http://www.muckrock.com/foi/miami-103/miami-pd-hunchlab-documents-30109/ [http://perma.cc/9WPG-YH3F] (displaying correspondence).

168 *See St. Louis County PD HunchLab Documents*, MUCKROCK, http://www.muckrock.com/foi/st-louis-county-8838/st-louis-county-pd-hunchlab-documents-30113/ [http://perma.cc/P68R-8M4F] (displaying the exchange of correspondence).

169 *Purchase Order*, CITY OF PHILADELPHIA PROCUREMENT DEP'T. (2015) http://cdn.muckrock.com/foia_files/2017/02/06/POXX16106457_-_Redacted.pdf [http://perma.cc/FQ6W-TEDN].

170 *See* Email from Robert Kieffer, Assistant City Solicitor to Michael Morisy (Feb. 6 2017), http://cdn.muckrock.com/foia_files/2017/02/06/Final_Response_-_Morisy_CP_2016-1710.pdf [http://perma.cc/46LH-M3A2].

window of police incident reports.[171] 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.[172]

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 "gradient boosting decision tree" 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.[173] A police officer cannot know how the algorithm's decision making 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 a new "modeling run" every few weeks to re-calibrate the model, and each of those modeling runs creates a new predictive algorithm.[174]

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."[175] 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

---

171 *See* Email from Tonya Peters, Police Legal Advisor, Lincoln Police Dep't to Michael Morisy (Nov. 30 2016), http://cdn.muckrock.com/foia_files/2016/11/30/PRR_Response.pdf [http://perma.cc/2KU2-DP3Z].

172 Police Incident Reports, http://cdn.muckrock.com/foia_files/2016/11/30/Archive.zip [http://perma.cc/7D2N-EL2Q].

173 *See* HUNCHLAB, *A Citizen's Guide to HunchLab*, *supra* note 77, at 5-6.

174 *Id.* at 19.

175 *Id.* at 2.

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.[176]

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.[177] Second, HunchLab recommends that the algorithm incorporate randomness to assure that police are not assigned to the same routes every day, in order to combat monotony on the job and to reduce the negative side effects of constant police presence in an area.[178] 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, and the like, and monitoring the effectiveness of the tactics used over time.[179]

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.[180] In general, 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

---

176 *See id.* at 26.

177 *Id.* at 9-10.

178 *Id.* at 10-11.

179 *Id.* at 11-14.

180 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, 47 N.Y.S.3d 838 (N.Y. Sup. Ct. 2016). *See generally* Valerie Strauss, *Judge Calls Evaluation of N.Y. Teacher 'Arbitrary' and 'Capricious' in Case Against New U.S. Secretary of Education*, WASH. POST (2016) (reporting on teacher's litigation against the city for wrongful termination), http://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/ [http://perma.cc/5CH4-RYQE].

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.[181] 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.[182] The New York State Education Department produced a number of documents, including the original contract with its vendor, the American Institutes for 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 focus on the New York VAM implementation; and sample outputs of the VAM algorithm—outputs for fifty students and fifty 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, fifty 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 Institutes for 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."[183] Thus, the algorithm or algorithms are not publicly available, and the process by which they were constructed has not been disclosed.

C. *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

---

181 Ours was not the first attempt. Cathy O'Neil also tried and failed to obtain New York City's VAM records. *See* Cathy O'Neill, *An Attempt To FOIL Request the Source Code of the Value-Added Model*, MATHBABE (2014), http://mathbabe.org/2014/03/07/an-attempt-to-foil-request-the-source-code-of-the-value-added-model/ [http://perma.cc/HNB3-LC5X].

182 *See New York State or New York City Value Added Measures for Teachers*, MUCKROCK (November 9, 2016), http://www.muckrock.com/foi/new-york-city-17/new-york-state-or-new-york-city-value-added-measures-for-teachers-29739/ [http://perma.cc/MN4R-U3U5] (displaying correspondence).

183 *See* Contract No. C010834 between the People of the State of New York and Am. Insts. for Research (Sept. 19, 2011), app. D, http://robertbrauneis.net/algorithms/AIR_CONTRACT_2011_Redacted_FOIL_final.pdf [http://perma.cc/6SP3-2Y74].

claimed to be either generally exempt from open records acts (as, for example, courts) or unable to comply with requests because they had promised to keep information confidential. 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 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.

IV. PRINCIPAL 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 important records or to deliver those records to the government 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.[184] Our research suggested that governments simply did not have many records concerning the creation and implementation of algorithms, either because those records were never generated or because they 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 any record of what problems the models were supposed to address, and what the metrics of success were.

---

[184] *See, e.g.*, *A Pocket Guide to the California Public Records Act*, FIRST AMEND. PROJECT SOC'Y PROF. JOURNALISTS, http://www.thefirstamendment.org/publicrecordsact.pdf [http://perma.cc/K3KP-Y8QT] ("The PRA covers only records that already exist, and an agency cannot be required to create a record, list, or compilation.").

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, justified, or recorded. In the best cases, there will be public requests for proposals for private vendors to supply predictive algorithms to government.[185] More typically, there will simply be a form agreement with a private 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[186] and assert trade secret protection over the algorithm and associated development and deployment processes.[187] 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.

---

185 *See, e.g.*, *Request for Proposal to Design and Implement Decision Support Tools and Predictive Analytics in Human Services*, ALLEGHENY COUNTY DEP'T HUM. SERV. (2014), http://www.county.allegheny.pa.us/Human-Services/Resources/Doing-Business/Solicitations/2014/Decision-Support-Tools-and-Predictive-Analytics-in-Human-Services-RFP.aspx [http://perma.cc/UHS3-8VW3].

186 *See* Joh, *supra* note 163, at 7 (discussing police department nondisclosure agreements with the Harris Corporation for use of Stingray police surveillance technology).

187 *See e.g.*, Amended Summary Judgment Op., Hous. Fed'n of Teachers v. Hous. Indep. Sch. Dist., Civil Action No. H-14-1189, at *1 (S.D. Tex. May 4, 2017) ("[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."). *See generally* Kitchin, *Thinking Critically*, *supra* note 6, at 20 ("[I]t is often a company's algorithms that provide it with a competitive advantage and they are reluctant to expose their intellectual property even with non-disclosure agreements in place").

The trade secret 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.[188] In litigation related to the algorithm, the Wisconsin Supreme 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."[189]

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.[190] It was also true of Alaska, using the Eckerd Rapid Safety Feedback risk assessment for children.[191] The open records laws of Arizona,[192] California,[193] and Alaska[194] all exempt trade secrets and confidential information, and none entitles public access to records the government does not have. It would require considerable additional 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.[195]

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

---

188 *See* Angwin et al., *supra* note 7; *see also* Diakopolous, *supra* note 9 (discussing trade secrecy barriers to disclosure).

189 State v. Loomis, 881 N.W.2d 749, 763-64 (Wis. 2016). It also required disclosure that no validation studies have been completed and tools must be constantly monitored and re-normed. *Id.* at 769.

190 *See supra* text accompanying notes 129-130.

191 *See* Illinois FY16 Contract, *supra* note 145; Illinois FY17 Contract, *supra* note 145.

192 Arizona Public Records Law, ARIZ. REV. STAT. ANN. § 39-121 (2017).

193 California Public Records Act, CAL. GOVT. CODE §§ 6250-6276.48 (West 2017).

194 Alaska Public Records Disclosures, ALASKA STAT. §§ 40.25.100-350 (West 2017).

195 As noted above, Florida's Seventh Judicial District took a step in the right direction, *see supra* text accompanying note 129, but it could have done much more.

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.[196] 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.[197] 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.

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.[198] Most states have adopted the Uniform Trade Secret Act,[199] 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

---

196 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 [http://perma.cc/BNX9-ZK8R].

197 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." Customer Success Training Bulletin, SST (2015), http://www.documentcloud.org/documents/3221020-ShotSpotter-nationwide-memo-July-2015.html [http://perma.cc/F2DB-8AN9]; *see also* Jason Tashea, *Should the Public Have Access to Data Police Acquire Through Private Companies?*, ABA J. (Dec. 1, 2016), http://www.abajournal.com/magazine/article/public_access_police_data_private_company [http://perma.cc/B3L2-DYMK] (reporting on municipal discomfort with Shotspotter assertions of ownership of data it collects on gunshots fired within the jurisdiction).

198 *See, e.g.*, WASH. REV. CODE § 42.56.060 (2017) ("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 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).

199 For a table of jurisdictions adopting the act, see UNIF. TRADE SECRETS ACT (UNIF. LAW COMM'N 1985) (West 2016).

secret was . . . acquired under circumstances giving rise to a duty to maintain its secrecy or limit its use."[200] Governments are persons, and therefore potentially liable.[201] Thus, vendors create a pull towards secrecy by asserting protection or demanding that government officials sign nondisclosure agreements. It is a pull strengthened by the government agency's own interests in secrecy, for reasons discussed below.

Open records acts do not exert as much of a counterforce for transparency as they might, because they generally exempt trade secrets.[202] Exemption 4 of FOIA 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."[203] The exemption thus covers two broad categories: (1) trade secrets; and (2) information that is (a) commercial or financial, (b) obtained from a person, and (c) privileged or confidential.

Under scrutiny, these trade secret exemptions are narrower than companies might claim. Overly generous agency

---

200 *Id.* § (1)(2)(ii).

201 *Id.* § (1)(3). In addition, federal law specifically forbids disclosure of trade secrets. 18 U.S.C. § 1905 (imposing criminal liability on any U.S. government employee who in the 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").

202 *See, e.g.*, CONN. GEN. STAT. § 1-210(b)(5) (2017) (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) (2017) (deeming not to be public "[t]rade secrets and commercial or financial information obtained from a person which is of a privileged or confidential nature"); Pennsylvania Right-to-Know Law, 65 PA. STAT. AND CONS. STAT ANN. § 67.708 (b)(11) (exempting from disclosure "[a] record that constitutes or reveals a trade secret or confidential proprietary information"). In other states, courts recognize trade secrets under more general exemptions. *See, e.g.*, Phx. Newspapers, Inc. v. Keegan, 35 P.3d 105, 112 (Ariz. Ct. App. 2001) (recognizing that trade secrets are "protected by the confidentiality exception to disclosure" in Arizona's open records law). *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) (discussing the coverage of state trade secrets exemptions from open records laws); *Open Government Guide*, REPORTERS COMMITTEE FOR FREEDOM PRESS, http://www.rcfp.org/ogg/index.php [http://perma.cc/2QNF-V83V] (linking to all fifty state open records acts).

203 5 U.S.C. § 552(b)(4) (2016).

protections have been struck down when challenged. 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[204] (and the federal Defend Trade Secrets Act of 2016).[205] 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."[206] There must be "a direct relationship between the information at issue and the productive process," rather than merely "collateral business confidentiality."[207] 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 trade secrets in the commercial sphere should count as trade secrets for open records exemption purposes.

The second prong of Exemption 4 permits secrecy for some kinds of financial or commercial information.[208] This part of the exemption is also limited. The information has to be "privileged or confidential."[209] The D.C. Circuit has held that a mere promise of confidentiality to the source of the information is insufficient. Rather, the government must prove with respect to compelled records that disclosure would likely (1) "impair the government's

---

204 UNIF. TRADE SECRETS ACT § 1.4 (defining a trade secret as "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")

205 18 U.S.C. § 1839(3) (2016) (including "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").

206 Pub. Citizen Health Research Grp. v. U.S. Food & Drug Admin., 704 F.2d 1280, 1288-89 n.25 (D.C. Cir. 1983) ("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).

207 *Pub. Citizen*, 704 F.2d at 1287-88.

208 *Guide to the Freedom Of Information Act*, U.S. DEP'T JUSTICE (May 2004), http://www.justice.gov/oip/foia-guide-2004-edition-exemption-4 [http://perma.cc/LC5X-EJPK] (stating that records are commercial if the submitter "has a 'commercial interest' in them").

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

ability to obtain necessary information in the future" or (2) "cause substantial harm to the competitive position" of the information source.[210] With respect to voluntarily disclosed records, commercial and financial information is "confidential" if the source would not customarily release such information to the public.[211] In sum, the FOIA trade secret exemption applies only to a narrow range of "bet the company" trade secrets and a subset of financial or commercial information. The presumption of disclosure remains. 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.[212]

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,[213] or in an alternative formulation, to "encourage individuals to provide certain kinds of confidential information to the Government."[214] Similarly, state open records laws protect trade secrets and confidential information *in order to* advance public goals.[215] 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 a government is acting as a customer and not as a regulator, because secrecy is not abetting the government's regulatory power. Currently pending litigation in New York poses the question of how far trade secret claims should be honored when the government acts

---

210 Nat'l Parks & Conservation Ass'n v. Morton, 498 F.2d 765, 770 (D.C. Cir. 1974).

211 Critical Mass Energy Project v. Nuclear Regulatory Comm'n, 975 F.2d 871, 872-74 (D.C. Cir. 1992).

212 *See, e.g.*, Pennsylvania Right-to-Know Law, 65 PA. STAT. AND CONS. STAT ANN. § 67.707 (providing that a 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 ten days of notifying the company, the state agency must decide to release or withhold the information).

213 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." *Guide to the Freedom Of Information Act*, *supra* note 208; *see also id.* at n.2.

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

215 *See, e.g.*, Verizon N.Y., Inc. v. N.Y. State Pub. Serv. Comm'n, 23 N.Y.S.3d 446 (N.Y. App. Div. 2016) ("[T]he 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.").

in its enterprise capacity. Citing trade secret protection,[216] New York City refused the Brennan Center for Justice's freedom of information law requests for records related to the sentencing algorithm known as Palantir Gotham.[217]

All of the requests we made were submitted to jurisdictions acting in their enterprise capacities. 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.[218] 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."[219] 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.

---

216 N.Y. PUB. OFF. L. § 87(2)(d) (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")

217 Memorandum of Law in Support of Verified Petition, Brennan Center for Justice at New York University School of Law v. New York City Police Department, Case No. 0160541/2016 (N.Y. Sup. Ct. 2016) (Doc. No. 8), http://www.brennancenter.org/sites/default/files/8%20-%20Memorandum%20of%20Law%20in%20Support%20of%20Verified%20Petition.pdf [http://perma.cc/E5E6-Q7HB].

218 Kroll et al., *supra* note 26, at 15-17; *see also* David S. Levine, *Secrecy and Unaccountability: Trade Secrets in Our Public Infrastructure*, 59 FLA. L. REV. 135, 180-81 (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).

219 *See* Levine, *supra* note 218, at 157; *see also* Mark Fenster, *The Opacity of Transparency*, 91 IOWA L. REV. 885, 918-19 (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).

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.[220] If a criminal defendant knows that statements she makes will result in a higher recidivism risk score, she may lie.[221] 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.[222]

These concerns are understandable, but do not excuse non-responsiveness to open records requests. Open records acts do address potential gaming in the context of law enforcement investigations and investigative techniques.[223] Exemption 7(E) of FOIA asks explicitly whether the disclosure of investigative

---

220 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 KNOWLEDGE DISCOVERY AND DATA MINING (KDD) 641 (Robert Grossman et al. eds., 2005); Pavel Liskov & Richard Lippmann, *Machine Learning in Adversarial Environments*, 81 MACHINE LEARNING 115 (2010).

221 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 et al., *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., 2016).

222 *See* JEROME P. BJELOPERA, BART ELIAS & AARON SISKIN, CONG. RESEARCH SERV., R44678, THE TERRORIST SCREENING DATABASE AND PREVENTING TERRORIST TRAVEL 12 (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 58, at 1210 (noting that 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).

223 *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").

techniques would "risk circumvention of the law."[224] However, at its core, "investigation" concerns the identification of perpetrators and gathering of evidence of crimes that have already been committed. In exceptional cases, some courts have been willing to stretch "investigation" to cover some preventative measures,[225] and one of our open records requests revealed that one jurisdiction exempted itself from providing data related to police surveillance techniques, arguably a cousin of prevention.[226] Predictive policing programs like PredPol and HunchLab, however, which are focused on deterrence rather than investigation, 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.[227]

Agencies may best deal with concerns about gaming by adopting algorithms that are relatively immune to manipulation. For example, the Arnold Foundation claims that

---

224 5 U.S.C. § 552(b)(7)(E).

225 *See* Coastal Delivery Corp. v. U.S. Customs Serv., 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. U.S. Bureau of Reclamation, 272 F. Supp. 2d 1313, 1320-22 (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).

226 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* Legislation Details (With Text), File # 15-361, *supra* note 164; FLA. STAT. § 119.071(2)(d) (2017) ("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.

227 As we mentioned above, *see supra* page 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.

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.[228] 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.[229]

Another concern officials might have is that they do not 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.[230] None of our open records requests were 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 that agencies have already announced a rule and explained its rationale. The point of exempting the deliberative 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."[231] 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.

The judicial branch is often exempt from open records laws.[232] 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 in every case, but it

---

228 *See Developing a National Model for Pretrial Risk Assessment*, LAURA & JOHN ARNOLD FOUND. (Nov. 2013), http://www.arnoldfoundation.org/wp-content/uploads/2014/02/LJAF-research-summary_PSA-Court_4_1.pdf [http://perma.cc/KTY8-VCP9] (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").

229 *See A Citizen's Guide to HunchLab*, *supra* note 77, at 10-11.

230 *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"); 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"); N.Y. PUB. OFF. L. 87(2)(g) (exempting most "inter-agency or intra-agency materials").

231 Coastal States Gas Corp. v. Dep't of Energy, 617 F.2d 854, 866 (D.C. Cir. 1980).

232 *See, e.g.*, 5 U.S.C. § 552(f)(1) (defining "agency" to exclude courts); 65 PA. STAT. & CONS. STAT. ANN. § 67.304 (West 2017) (requiring "[j]udicial agencies" only to provide access to financial records).

should be. The formulation and adoption of an algorithm for a court system bears little resemblance to judicial decision making 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,"[233] and the Supreme Court also uses notice-and-comment rulemaking under procedures issued by the Judicial Conference.[234] State courts have similar public procedures.[235] 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.

V. 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.[236] Even if that were to happen, removal of trade

---

233 28 U.S.C. § 2071(b).

234 *See* ADMIN. OFFICE OF THE U.S. COURTS, PROCEDURES FOR COMMITTEES ON RULES OF PRACTICE AND PROCEDURE § 440.20.40, http://www.uscourts.gov/rules-policies/about-rulemaking-process/laws-and-procedures-governing-work-rules-committees-0 [http://perma.cc/MEL6-GJ3A].

235 *See, e.g.*, ILL. S. CT. R. 3(a)(1) (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").

236 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*, 70 STAN. L. REV. (forthcoming 2018).

secret protection would not itself solve the problem of inadequate documentation and private possession of records. A more fruitful course would be for governments to use their contracting powers to insist on appropriate record creation, provision, and disclosure.[237] 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.[238] 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.[239] 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 over-claiming trade secrets, and forces the contractor to consider exactly why and how the disclosure of particular information would undermine its competitive position.[240] Such language dovetails with appropriately narrow construction of trade secret exemptions in open records acts.[241]

---

237 *Cf.* Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553, 589-90 (1998) (arguing for the use of public procurement standards to pursue policy goals).

238 *See supra* notes 127-128.

239 *See supra* note 129.

240 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 for Research, *supra* note 183.

241 *See* text accompanying notes 204-217 (discussing appropriately narrow construction of trade secret exemptions in open records acts).

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 impacting how governments run. Thus, governments must understand that they have leverage even if they are not paying for services.[242]

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[243]—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.[244] By contrast, when the Alaska Department of Health and Social Services signed a memorandum of understanding 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 Rapid Safety Feedback program, including all software and all reports that the software produced.[245]

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

---

242 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 decision-making 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.

243 *See* Illinois FY16 Contract, *supra* note 145, at 11.

244 *See id.* at 11-12.

245 *See* Memorandum of Understanding of February 20, 2015 between Eckerd Youth Alternatives, Inc. and the Alaska Department of Health and Social Services, *supra* note 142.

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 on a license for its own use and use by other jurisdictions. Thus, for example, Allegheny County's contract with the Auckland Consortium grants a nonexclusive 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.[246] 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,[247] while the Alaska agreement cedes ownership of all reports to Eckerd.[248]

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 the 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 (coming into force in 2018), which stipulates that the

---

246 *See* AUT Enterprises Ltd. Contract 9-1-14 to 6-30-15, at 37-39, 45-46 (2014) http://robertbrauneis.net/algorithms/AlleghenyAUTContract.pdf [http://perma.cc/5HWR-CR8U].

247 *See* Illinois FY16 Contract, *supra* note 145, § 4.8 S b.

248 *See* Memorandum of Understanding of February 20, 2015 between Eckerd Youth Alternatives, Inc. and the Alaska Department of Health and Social Services, *supra* note 142.

function of an algorithm must be made understandable to the public.[249]

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.[250] The National Federation of Municipal Analysts has promulgated a series of best disclosure practices in connection with the issuance of municipal debt.[251] The Online Trust Alliance has released a number of best practices documents, including the Internet of Things Trust Framework 2.5, a set of privacy and security principles focused on connected home and wearable technologies.[252] 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."[253]

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 and application; relevant, available, and collectable data; considered exclusion of data;

---

249 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 O.J. (L 119/1) 71; *cf.* Sandra Wachter et al., *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIVACY L. 76 (2016) (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 decision making); Goodman & Flaxman, *supra* note 52, at 6 (identifying developer secrecy, public technical illiteracy, and algorithmic design as barriers to explanation).

250 *Opening Government: A Guide to Best Practice in Transparency, Accountability and Civic Engagement Across the Public Sector*, TRANSPARENCY & ACCOUNTABILITY INITIATIVE (2011), http://www.transparency-initiative.org/archive/wp-content/uploads/2011/09/15-Open-government11.pdf [http://perma.cc/V244-AJME].

251 *Disclosure Guidelines*, NAT'L FED'N MUN. ANALYSTS, http://www.nfma.org/disclosure-guidelines [http://perma.cc/DC9F-WDAV].

252 IoT Security & Privacy Trust Framework v. 2.5, ONLINE TRUST ALLIANCE (2017), http://otalliance.actonsoftware.com/acton/attachment/6361/f-008d/1/-/-/-/-/IoT%20Trust%20Framework.pdf [http://perma.cc/LB23-EM4N].

253 *Statement on Algorithmic Transparency and Accountability*, ASS'N FOR COMPUTING MACH. U.S. PUB. POLICY COUNCIL, (Jan. 12, 2017), http://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf [http://perma.cc/5TKE-LFHT].

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 and Application

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 more narrowly, 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 to the fairness of its parole decision practices. Each of these situations will likely call for different sensitivities in creating predictive algorithms. Predictive algorithms can also be applied to assist governmental decisions at a variety of junctures. For example, while the Allegheny Family Screening Tool and Eckerd Rapid Safety Feedback both provide child welfare assessments, the former is designed to be applied at the moment a call comes in to a child welfare hotline, as an immediate screening tool; the latter is apparently used to periodically review all child welfare cases currently being handled by an agency. There should ideally be some reflection on the particular decision-making process for which an algorithm is being designed, whether that the best application of algorithmic prediction in the operations of that agency, and whether the algorithm design is appropriate for that application.

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. Documenting all of these is important for understanding the training data and input data of the algorithm.

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.[254] 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

---

254 *Cf.* Julia Andre, Luis Ceferino & Thomas Trinelle, *Prediction Algorithm for Crime Recidivism*, MACH. LEARNING PROJECT, STAN. U. 1 (2015), http://cs229.stanford.edu/proj2015/250_report.pdf [http://perma.cc/YWP2-5CTX] (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").

facts gathered in an interview with the criminal defendant.[255] 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.[256]

In the case of HunchLab, the Lincoln Police Department revealed that the output that HunchLab produces on any given day is based on police incident reports for the previous thirty days.[257] The choice of a thirty-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 three

---

255 *Developing a National Model for Pretrial Risk Assessment*, *supra* note 228, at 3.

256 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), http://crimemapping.info/wp-content/uploads/2015/03/CrimeMappingNews_ Issue23.pdf [http://perma.cc/DF5F-FBFF].

257 *See* Email from Tonya Peters, *supra* note 171.

hundred different jurisdictions nationwide.[258] 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."[259] 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 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. Therefore, the choices of criteria used to predict should be documented, especially when they diverge from the obvious.

For example, the general predictive goal of an algorithm may be to predict where and when crime will occur, but the only available training and input data are most likely crimes that

---

[258] *Developing a National Model for Pretrial Risk Assessment*, *supra* note 228 at 3.

[259] *The Public Safety Assessment (PSA)*, LAURA & JOHN ARNOLD FOUND., http://www.arnoldfoundation.org/wp-content/uploads/PSA-Infographic.pdf [http://perma.cc/PLR5-2SZ6].

have been reported, and that have been reported relatively soon after their occurrence. Thus, the algorithm will end up more specifically predicting not where crimes will occur, but where crimes *that will be reported* will occur. That is troubling, not just because many crimes are not reported,[260] but because crimes are reported at different rates in different neighborhoods.[261] For example, one study found that simple assaults were less likely to be reported in disadvantaged neighborhoods.[262] Another found that crimes were particularly underreported in heavily immigrant neighborhoods.[263] 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.[264] Thus, an algorithm trained on reported crimes may end up directing police away from disadvantaged, immigrant, and young victims, who are arguably among the most vulnerable. 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 as the specific predictive criteria, for reasons explained at length in the Auckland Consortium report.[265] Similarly, the COMPAS recidivism algorithm is trained on data about repeat arrests for crimes, not data about convictions;[266] although the Arnold Foundation has not disclosed

---

260 *See* Lynn Langton et al., *Victimizations Not Reported to Police, 2006-2010*, U.S. DEP'T JUST., BUREAU JUST. STAT. (Aug. 2012), http://www.bjs.gov/content/pub/pdf/vnrp0610.pdf [http://perma.cc/ASK6-TJ82].

261 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*, UPTURN 5 (2016), www.teamupturn.org/static/reports/2016/stuck-in-a-pattern/files/Upturn_-_Stuck_In_a_Pattern_v.1.01.pdf [http://perma.cc/C5SL-MF4Q].

262 Eric P. Baumer, *Neighborhood Disadvantage and Police Notification by Victims of Violence*, 40 CRIMINOLOGY 579, 597 (2002).

263 Carmen M. Gutierrez & David S. Kirk, *Silence Speaks: The Relationship Between Immigration and the Underreporting of Crime*, 63 CRIME & DELINQUENCY 928, 946 (2015)

264 *See* Stacey J. Bosick et al., *Reporting Violence to the Police: Predictors Through the Life Course*, 40 J. CRIM. JUST. 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, *supra* note 77, at 25-26.

265 *See* Vaithianathan et al., *supra* note 72.

266 *See* COMPAS Risk & Need Assessment System, NORTHPOINTE 2 (2012), http://www.northpointeinc.com/files/downloads/FAQ_Document.pdf [http://perma.cc/6TL7-GCA4] (describing the training method for the General

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?"[267] Arrests of minority community members will be skewed artificially high, and therefore other predictive criteria had to be chosen; those choices should be disclosed.

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.[268] 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. It may be the case that a non-linear predictive model (for example, because it uses cutoffs of particular ages) produces comparably statistically significant results that are just as statistically significant.

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.[269] Documentation of those development techniques is also likely a best practice.

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

---

Recidivism Risk Scale algorithm as based on data on whether defendants have been arrested within two years of an intake assessment).

267 Abe Gong, *Ethics of Powerful Algorithms (2 of 4)*, MEDIUM (July 12, 2016), http://medium.com/@AbeGong/ethics-for-powerful-algorithms-2-of-3-5bf750ce4c54 [http://perma.cc/VVR9-PF9G].

268 *See, e.g.*, SHAI SHALEV-SCHWARTZ & SHAI BEN-DAVID, UNDERSTANDING MACHINE LEARNING: FROM THEORY TO ALGORITHMS 89-240 (2014).

269 *See, e.g.*, YASER S. ABU-MOSTAFA, MALIK MAGDON-ISMAIL & HSUAN-TIEN LIN, LEARNING FROM DATA: A SHORT COURSE 138-54 (2012).

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.

Audits could serve as alternatives or additions to validation studies. Where optimal disclosure will not happen for trade secret, security, or privacy reasons, it could be important to have a third-party confidential audit of algorithm development.[270] 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.[271] Public entities should also contract for audits of algorithm implementation, which is what the Seventh Judicial Circuit got for its implementation of the PSA algorithm (performed by an Arnold Foundation subcontractor).[272] 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

---

[270] On algorithm audits, see Christian Sandvig et al., *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms*, ANN. MEETING INT'L COMM. ASS'N (2014) http://tiny.cc/61wrmy [http://perma.cc/F29K-BALH].

[271] *See Statement on Algorithmic Transparency and Accountability*, *supra* note 253, at 2 ("Auditability: Models, algorithms, data, and decisions should be recorded so that they can be audited in cases where harm is suspected.").

[272] *See* Dal Pra, *Volusia County Case Review*, *supra* note 132.

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.[273] This will allow both for public accountability and for the users of the algorithm to judge its output. If the algorithm is so complicated that a plain-language explanation does not seem possible, that should probably be 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 government agents (or people they trust) understand the algorithm, they will be better equipped to accept its judgment or override it.[274]

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.[275]

VI. CONCLUSION

There will always be value in public entities using open source code, or otherwise releasing the code running predictive analytics. But access to code will not usually be necessary to achieve meaningful transparency, and sometimes will not even

---

273 *See id.* ("Explanation: Systems and institutions that use algorithmic decision-making are encouraged to produce explanations regarding both the procedures followed by the algorithm and the specific decisions that are made. This is particularly important in public policy contexts."), Diakopoulos, *supra* note 6, at 411 (recommending that a transparency policy for algorithms include "the definitions, operationalizations, or thresholds used by similarity or classification algorithms").

274 On the development of interpretable algorithms, see Jiaming Zeng, Berk Ustun & Cynthia Rudin, *Interpretable Classification Models for Recidivism Prediction*, J. ROYAL STAT. SOC'Y: SERIES A (STATISTICS IN SOC'Y) (2016), http://arxiv.org/abs/1503.07810 [http://perma.cc/MZ87-LNF9].

275 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 MED. 1246 (2012).

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 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 contractors' competitive positions. To the extent that irreducible trade secrets remain in predictive algorithm projects, government records custodians responding to open records requests should construe those claims narrowly. Courts should do the same, requiring contractors to release records (even in redacted form) that will not weaken their competitive position. This will allow for meaningful transparency, and thus government accountability in the use of these algorithms.