

**ALGORITHMIC IMPACT ASSESSMENTS:**  
A PRACTICAL FRAMEWORK FOR PUBLIC AGENCY  
ACCOUNTABILITY

Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittaker

APRIL 2018

AINOW

# TABLE OF CONTENTS

<b>Executive Summary</b>	<b>3</b>
<b>I. The Algorithmic Impact Assessment Process</b>	<b>7</b>
A. Pre-acquisition review	8
B. Initial agency disclosure requirements	9
C. Comment period	9
D. Due process challenge period	10
E. Renewing AIAs	10
<b>II. The Content of an Algorithmic Impact Assessment</b>	<b>11</b>
A. Establishing Scope: Define “automated decision system”	11
Challenge: Drawing boundaries around systems	
Definitions and the EU General Data Protection Regulation	
B. Public notice of existing and proposed automated decision systems: Alert communities about the systems that may affect their lives	13
Challenge: Trade secrecy	
C. Internal agency self-assessments: Increase the capacity of public agencies to assess fairness, justice, due process, and disparate impact	15
Opportunity: Benefit to vendors	
Opportunity: AIAs and public records requests	
Challenge: Considering both allocative and representational harms	
D. Meaningful access: Allow researchers and auditors to review systems once they are deployed	18
Challenge: Funding and resources	
<b>III. Conclusion</b>	<b>21</b>
<b>Acknowledgements</b>	<b>22</b>



# EXECUTIVE SUMMARY

## Public agencies urgently need a practical framework to assess automated decision systems and to ensure public accountability

Automated decision systems are currently being used by public agencies, reshaping how criminal justice systems work via risk assessment algorithms<sup>1</sup> and predictive policing,<sup>2</sup> optimizing energy use in critical infrastructure through AI-driven resource allocation,<sup>3</sup> and changing our employment<sup>4</sup> and educational systems through automated evaluation tools<sup>5</sup> and matching algorithms.<sup>6</sup>

Researchers, advocates, and policymakers are debating when and where automated decision systems are appropriate, including whether they are appropriate at all in particularly sensitive domains.<sup>7</sup> Questions are being raised about how to fully assess the short and long term impacts of these systems, whose interests they serve, and if they are sufficiently sophisticated to contend with complex social and historical contexts. These questions are essential, and developing strong answers has been hampered in part by a lack of information and access to the systems under deliberation. Many such systems operate as “black boxes” – opaque software tools working outside the scope of meaningful scrutiny and accountability.<sup>8</sup> This is concerning, since an informed policy debate is impossible without the ability to understand which existing systems are being used, how they are employed, and whether these systems cause unintended

- 1 Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, “Machine Bias,” *ProPublica*, May 23, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- 2 Jack Smith IV, “Crime-prediction tool PredPol amplifies racially biased policing, study shows,” *Mic*, Oct. 9, 2016, <https://mic.com/articles/156286/crime-prediction-tool-pred-pol-only-amplifies-racially-biased-policing-study-shows#.DZeqQ4LYs>; Andrew G. Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, (New York: NYU Press, 2017).
- 3 James Vincent, “Google uses DeepMind AI to cut data center energy bills,” *The Verge*, July 21, 2016, <https://www.theverge.com/2016/7/21/12246258/google-deepmind-ai-data-center-cooling>.
- 4 Stephen Buranyi, “‘Dehumanising, impenetrable, frustrating’: the grim reality of job hunting in the age of AI,” *The Guardian*, March 4, 2018, <https://www.theguardian.com/inequality/2018/mar/04/dehumanising-impenetrable-frustrating-the-grim-reality-of-job-hunting-in-the-age-of-ai>.
- 5 Laura Moser, “A Controversial Teacher-Evaluation Method Is Heading to Court. Here’s Why That’s a Huge Deal,” *Slate*, Aug. 11, 2015, [http://www.slate.com/blogs/schooled/2015/08/11/vam-lawsuit\\_in\\_new\\_york\\_state\\_here\\_s\\_why\\_the\\_entire\\_education\\_reform\\_movement.html](http://www.slate.com/blogs/schooled/2015/08/11/vam-lawsuit_in_new_york_state_here_s_why_the_entire_education_reform_movement.html).
- 6 Benjamin Herold, “Custom Software Helps Cities Manage School Choice,” *Education Week*, March 18, 2018, [https://www.edweek.org/ew/articles/2013/12/04/13algorithm\\_ep.h33.html](https://www.edweek.org/ew/articles/2013/12/04/13algorithm_ep.h33.html).
- 7 See for example, Kade Crockford, “Risk assessment tools in the criminal justice system: inaccurate, unfair, and unjust?,” *ACLU of Massachusetts*, March 8, 2018, <https://privacysos.org/blog/risk-assessment-tools-criminal-justice-system-inaccurate-unfair-unjust>; Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, (New York: St. Martin’s Press, 2018); Nazgol Ghandnoosh, *Black Lives Matter: Eliminating Racial Inequity in the Criminal Justice System* (Washington DC: The Sentencing Project, 2015), <http://sentencingproject.org/wp-content/uploads/2015/11/Black-Lives-Matter.pdf>; Insha Rahman, “The State of Bail: A Breakthrough Year for Bail Reform,” *Vera Institute of Justice*, 2017, <https://www.vera.org/state-of-justice-reform/2017/bail-pretrial>.
- 8 Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press, 2015).

consequences. The Algorithmic Impact Assessment (AIA) framework proposed in this report is designed to support affected communities and stakeholders as they seek to assess the claims made about these systems, and to determine where – or if – their use is acceptable.

It is not simply affected communities who lack the necessary information to assess how automated decision systems are working. Governments themselves are also struggling to assess how these systems are used, whether they are producing disparate impacts, and how to hold them accountable. Currently, few agencies are explicitly mandated to disclose anything about the systems they have in place or are planning to use.<sup>9</sup> Instead, impacted communities, the public at large, and governments are left to rely on what journalists, researchers, and public records requests have been able to expose.<sup>10</sup>

### KEY ELEMENTS OF A PUBLIC AGENCY ALGORITHMIC IMPACT ASSESSMENT

1. Agencies should conduct a self-assessment of existing and proposed automated decision systems, evaluating potential impacts on fairness, justice, bias, or other concerns across affected communities.
2. Agencies should develop meaningful external researcher review processes to discover, measure, or track impacts over time;
3. Agencies should provide notice to the public disclosing their definition of “automated decision system,” existing and proposed systems, and any related self-assessments and researcher review processes before the system has been acquired;
4. Agencies should solicit public comments to clarify concerns and answer outstanding questions; and
5. Governments should provide enhanced due process mechanisms for affected individuals or communities to challenge inadequate assessments or unfair, biased, or otherwise harmful system uses that agencies have failed to mitigate or correct.

If governments deploy systems on human populations without frameworks for accountability, they risk losing touch with how decisions have been made, thus making it difficult for them to identify or respond to bias, errors, or other problems. The public will have less insight into how agencies function, and have less power to question or appeal decisions. The urgency of this concern is why the AI Now Institute has called for an end to

<sup>9</sup> Catherine Crump, “Surveillance Policy Making by Procurement,” *Wash. L. Rev.* 91 (2016): 1595.

<sup>10</sup> Julia Angwin, et al., “Machine Bias”; Ali Winston, “Transparency Advocates Win Release of NYPD ‘Predictive Policing’ Documents,” *The Intercept*, Jan. 27, 2018, <https://theintercept.com/2018/01/27/nypd-predictive-policing-documents-lawsuit-crime-forecasting-brennan/>.

the use of unaudited “black box” systems in core public agencies.<sup>11</sup> The turn to automated decision-making and predictive systems must not prevent agencies from fulfilling their responsibility to protect basic democratic values, such as fairness, justice, and due process, and to guard against threats like illegal discrimination or deprivation of rights.

### **Implementing AIAs will help public agencies achieve four key policy goals**

AIAs will not solve all of the problems that automated decision systems might raise, but they do provide an important mechanism to inform the public and to engage policymakers and researchers in productive conversation. With this in mind, AIAs are designed to achieve four key policy goals:

1. Respect the public’s right to know which systems impact their lives by publicly listing and describing automated decision systems that significantly affect individuals and communities;
2. Increase public agencies’ internal expertise and capacity to evaluate the systems they build or procure, so they can anticipate issues that might raise concerns, such as disparate impacts or due process violations;
3. Ensure greater accountability of automated decision systems by providing a meaningful and ongoing opportunity for external researchers to review, audit, and assess these systems using methods that allow them to identify and detect problems; and
4. Ensure that the public has a meaningful opportunity to respond to and, if necessary, dispute the use of a given system or an agency’s approach to algorithmic accountability.

### **Algorithmic Impact Assessments offer a practical accountability framework combining agency review and public input**

Impact assessments are nothing new. We have seen them implemented in scientific and policy domains as wide-ranging as environmental protection,<sup>12</sup> human rights,<sup>13</sup> data protection,<sup>14</sup> and privacy.<sup>15</sup> AIAs draw on these frameworks and combine them with growing and important research that scientific and policy experts have been developing

---

11 AI Now 2017 Report, Recommendation #1, [https://ainowinstitute.org/AI\\_Now\\_2017\\_Report.pdf](https://ainowinstitute.org/AI_Now_2017_Report.pdf).

12 Leonard Ortolano and Anne Shepard, “Environmental impact assessment: challenges and opportunities,” *Impact assessment* 13, no. 1 (1995): 3-30. <https://www.tandfonline.com/doi/abs/10.1080/07349165.1995.9726076>.

13 United Nations, “Guiding Principles on Business and Human Rights: Implementing the United Nations ‘Protect, Respect and Remedy’ Framework,” 20-24 (2011), [http://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR\\_EN.pdf](http://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf).

14 “Data Protection Impact Assessments,” Information Commissioner’s Office, accessed March 16, 2018, <https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/>.

15 Kenneth A. Bamberger and Deirdre Mulligan, “Privacy Decision Making in Administrative Agencies,” *Chicago L. Rev.* 75(1):75 (2008), <https://www.truststc.org/pubs/258.html>.

## EXECUTIVE SUMMARY

on the topic of algorithmic accountability.<sup>16</sup> AIAs also complement similar domain-specific proposals for algorithmic accountability, like Andrew Selbst's recent work on Algorithmic Impact Statements in the context of predictive policing systems.<sup>17</sup> By integrating these approaches, AIAs can begin to shed light on automated decision systems, helping us better understand their use and determine where they are and are not appropriate, both before they are deployed and on a recurring basis when they are actively in use. While AIAs will not be a panacea for the problems raised by automated decision systems, they are designed to be practical tools to inform the policy debate about the use of such systems and to provide communities with information that can help determine whether those systems are appropriate.

---

16 See generally, Danielle Keats Citron, "Technological due process," *Wash. L. Rev.* 85 (2007): 1249; Lilian Edwards and Michael Veale, "Slave to the Algorithm? Why a 'Right to an Explanation' is Probably Not the Remedy You are Looking for," *16 Duke L. & Tech. Rev.* 18 (2017); Robert Brauneis and Ellen P. Goodman, "Algorithmic transparency for the smart city," *20 Yale J. L. & Tech.* 103 (2018); Danielle Keats Citron and Frank Pasquale, "The Scored Society: Due process for automated predictions," *Wash. L. Rev.* 89 (2014): 1; Andrew D. Selbst and Julia Powles, "Meaningful information and the right to explanation," *International Data Privacy Law* 7, no. 4 (2017): 233–242; Nicholas Diakopoulos, "Algorithmic Accountability: the investigation of Black Boxes," *Tow Center for Digital Journalism* (2014).; Solon Barocas and Andrew D. Selbst, "Big data's disparate impact," *Cal. L. Rev.* 104 (2016): 671; Kate Crawford and Jason Schultz, "Big Data and Due Process: Toward a framework to redress predictive privacy harms," *BCL Rev.* 55 (2014): 93.

17 Andrew D. Selbst, "Disparate Impact in Big Data Policing," *52 Georgia L. Rev.* 109 (2017), <https://ssrn.com/abstract=2819182>.

# I. THE ALGORITHMIC IMPACT ASSESSMENT PROCESS

Algorithmic Impact Assessments draw directly from impact assessment frameworks in environmental protection, data protection, privacy, and human rights policy domains.<sup>18</sup> For example, the United States’ National Environmental Protection Act mandates that federal agencies evaluate a proposed action’s impact on the “quality of the human environment” through an Environmental Impact Statement (EIS).<sup>19</sup> While the EIS process has by no means solved issues of environmental degradation, it has been credited with engendering increased sensitivity to environmental values within federal agencies and for informing the public, which is especially notable given the complex scientific knowledge the EIS process can require.<sup>20</sup>

The EIS process combines a focus on core values with a means for the public, outside experts, and policymakers to consider complex social and technical questions. As governments move to adopt new automated decision systems, AIAs can similarly help agencies and the public determine whether these systems promote fairness, justice, and due process or whether they infringe on those values.

In implementing AIAs, agencies should consider incorporating AIAs into the processes they already use to procure automated decision systems or any existing pre-acquisition assessment processes the agency already undertakes.<sup>21</sup> A pre-procurement AIA gives an agency the opportunity to engage the public and proactively identify concerns, establish expectations, and draw on expertise and understanding from relevant stakeholders.

While AIAs resemble environmental impact assessments, data protection impact assessments, or privacy impact assessments, they differ in some very important ways. For example, data protection impact assessments (DPIAs), like those mandated under Europe’s General Data Protection Regulation, similarly serve to highlight the data protection risks of automated systems used to evaluate people based on their personal data.<sup>22</sup> If a data controller finds a system to be “high risk,” then it must consult with its local governmental data protection authority.<sup>23</sup> However, DPIAs apply to both public and private organizations, are not shared with the public, and have no built-in external researcher review or other individualized due process mechanisms. AIAs, on the other hand, are explicitly designed to engage public agencies and the people they serve on these areas of concern through the various notice, comment, review, and due process elements. This allows a wide range of individuals, communities, researchers, and policymakers to participate in accountability efforts.

18 See *supra* notes 12-15.

19 42 U.S.C. § 4321, et seq.

20 Bamberger and Mulligan, “Privacy Decision Making”: “NEPA is now, however, considered by many in and out of agencies to have successfully ‘institutionaliz[ed] environmental values in government.’”

21 Crump, “Surveillance Policy Making by Procurement.”

22 “Data protection impact assessment,” Art. 35, Regulation (EU) 2016/679, of the European Parliament and 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.

23 Edwards and Veale, “Slave to the Algorithm?”

## A. PRE-ACQUISITION REVIEW

An Algorithmic Impact Assessment, much like an EIS, gives both the agency and the public the opportunity to evaluate the adoption of an automated decision system before the agency has committed to its use. **This allows the agency and the public to identify concerns that may need to be negotiated or otherwise addressed before a contract is signed.** This is also when the public and elected officials can push back against deployment before potential harms can occur.



A model procurement process as defined by the New York City Mayor's Office of Contract Services<sup>24</sup>

Of course, not all automated decision systems will come to an agency via standard procurement processes. There are many examples of systems acquired through in-kind donations, federal grants, and funding from private foundations. The data analysis company Palantir, for example, gave their analytics platform to the New Orleans Police Department pro bono, thus making it less visible to the New Orleans City Council.<sup>25</sup> The Seattle Police Department acquired a surveillance drone through a Department of Homeland Services grant program.<sup>26</sup> Similarly, the New York City and Los Angeles Police Departments have purchased

<sup>24</sup> New York City is in the process of evaluating algorithmic accountability as of 2018, so its typical processes provide a useful model for considering frameworks ("About Procurement," NYC Mayor's Office of Contract Services, accessed March 16, 2018, <https://www1.nyc.gov/site/mocs/about/procurement.page>).

<sup>25</sup> "License and Services Agreement," City of New Orleans & Palantir, dated Feb. 23, 2012, <https://www.documentcloud.org/documents/4344821-K12-168-Palantir-Technologies.html>.; Ali Winston, "Palantir has secretly been using New Orleans to test its predictive policing technology," *The Verge*, Feb. 27, 2018, <https://www.theverge.com/2018/2/27/17054740/palantir-predictive-policing-tool-new-orleans-nopd> (after the existence of Palantir's system came to light, the New Orleans Mayor's office declined to renew their contract with Palantir).

<sup>26</sup> Crump, "Surveillance Policy Making by Procurement."



## I. THE ALGORITHMIC IMPACT ASSESSMENT PROCESS

a variety of surveillance technologies using funding from local police foundations.<sup>27</sup> **An Algorithmic Impact Assessment should cover any automated decision system before it is deployed, no matter how it was acquired.**

### B. INITIAL AGENCY DISCLOSURE REQUIREMENTS

As part of the pre-acquisition review for an automated decision system, each agency will:

1. Publish their internal definition of “automated decision system”;
2. Publicly disclose information about each automated decision system, including details about its purpose, reach, potential internal use policies or practices, and implementation timeline;
3. Perform a self-assessment of each system, evaluate potential issues of inaccuracy, bias, and harms to affected communities, and establish ways to address these potential impacts, including proactive conversations or engagement with affected community members; and
4. Propose a plan for providing meaningful access to external researchers who seek to review the system once it is deployed.

The substance of each component is detailed in Section II. Depending on the particular acquisition schedule, capacity, and expertise of each agency, these disclosures could be made together or separately. Regardless, they must be made in the order above and all are required in order to fulfill the AIA policy goals. In the case of a pre-acquisition review, all disclosures should be made before the decision to use a given system is finalized.

### C. COMMENT PERIOD

The AIA process includes the opportunity for the public to engage with the agency over the content of its initial AIA disclosure. Agencies can decide how they want to organize the comment process: they could choose to separate each component of the AIA (“definition,” “disclosure,” “self-assessment,” and “meaningful access”) into separate comment periods or release the AIA as a single document and have one overarching comment period for that one document. There might be an advantage to agencies and the public in separating the definition of automated decision systems and the disclosure of systems before moving on to discuss internal assessments and external researcher access protocols. The initial disclosure provides a strong foundation for building public trust through appropriate levels of transparency, while subsequent requests can solicit further information or the presentation of new evidence, research, or other inputs that the agency may not have adequately considered.

---

27 Ali Winston, “NYPD Attempts to Block Surveillance Transparency Law with Misinformation,” *The Intercept*, July 7, 2017, <https://theintercept.com/2017/07/07/nypd-surveillance-post-act-lies-misinformation-transparency/>; “Private Donors Supply Spy Gear to Cops,” *ProPublica*, Oct. 13, 2014, <https://www.propublica.org/article/private-donors-supply-spy-gear-to-cops>.

### D. DUE PROCESS CHALLENGE PERIOD

The AIA process provides a much-needed basis for evaluating and improving agency systems. But without oversight, AIAs could become a checkbox that agencies mark off and forget, potentially sidelining community concerns.<sup>28</sup> That is why the Algorithmic Impact Assessment process should also provide a path for the public to challenge an agency if it fails to comply with AIA requirements or if its self-assessment process was deficient in adequately identifying or addressing key concerns. For example, if an agency fails to disclose a system that should have reasonably been considered an automated decision system, or if it allows vendors to make overbroad trade secret claims blocking meaningful system access,<sup>29</sup> the public should have the chance to raise concerns with an agency oversight body or directly in a court of law if the agency refuses to rectify these problems after the public comment period. The AIA process should give the public the opportunity to effectively challenge the agency's adoption of the system and prevent the system from being used when it fails to benefit affected communities.<sup>30</sup>

### E. RENEWING AIAs

In order to ensure their assessments remain current and incorporate the latest information and research, agencies should be required to renew AIAs on a regular schedule. The renewed AIA will also have renewed comment and due process challenge periods. For example, agencies could be required to conduct a new AIA on all of their systems every two years. However, if there have not been significant changes to the system, to the context of its deployment, or to the need for external research access, the agency should be allowed to minimally update their original AIA content as part of the renewal process.

---

28 The "checking the box" mentality is a common critique of workplace sexual harassment training (Yuki Noguchi, "Trainers, Lawyers Say Sexual Harassment Training Fails," *All Things Considered*, NPR, Nov. 8, 2017, <https://www.npr.org/2017/11/08/562641787/trainers-lawyers-say-sexual-harassment-training-fails>).

29 Rebecca Wexler, "Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System," *70 Stan. L. Rev.*, (forthcoming 2018), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2920883](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2920883); Natalie Ram, "Innovating Criminal Justice," *Northwestern L. Rev.* (forthcoming 2017), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3012162](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3012162).

30 In Santa Clara, California, for instance, a law passed in 2016 requires the local Board of Supervisors to explicitly approve new surveillance technology before moving forward with its use (Nicole A. Ozer, "Santa Clara County Passes Landmark Law to Shut Down Secret Surveillance," *ACLU of Northern California*, June 8, 2016, <https://www.aclunc.org/blog/santa-clara-county-passes-landmark-law-shut-down-secret-surveillance>).

## II. THE CONTENT OF AN ALGORITHMIC IMPACT ASSESSMENT

In line with the above process, the sections below outline each of the substantive aspects of an AIA, including various challenges that each agency would need to address.

### **A. ESTABLISHING SCOPE: DEFINE “AUTOMATED DECISION SYSTEM”**

In an AIA process, agencies must first publish their own definition of “automated decision system” that is both practical and appropriate for its particular context. This does not mean the agency must go through the effort of redefining “automated decision system” for each particular system: once they reach a working definition, they can choose to republish it in future AIAs as long as it continues to accurately describe the systems in ways that reinforce public trust and accountability. Agencies should also regularly revisit their definition when necessary to incorporate new types of systems, new applications of old systems, or research advances in relevant fields.

The flexibility of the AIA process allows an agency to publish their definition before conducting the rest of an AIA. This might be more efficient for the agency, so they can work out a definition before committing to a full review of a system that may not need it. Agencies could also save effort by borrowing definitions from other agencies and governments that are better tested, already have public approval, or perhaps have even withstood challenges in court. Agencies should be required to publish their definition at least once, even if they do not believe they have any “automated decision systems,” so that the public can evaluate the definition to see if it is reasonable.

This process of defining and specifying automated decision systems would help build agency capacity for the procurement and assessment of future systems, as experience with AIAs would help guide Requests for Proposals, budgeting, and other key milestones in the acquisition process.

### CHALLENGE: DRAWING BOUNDARIES AROUND SYSTEMS

Drawing an appropriate boundary around automated decision systems will be particular to each agency's context and the interests of the communities they serve. An overly-broad definition could burden agencies with disclosing systems that are not the main sources of concern. If a public servant uses a word processor to type up her notes from a meeting where some key decisions were made, and then checks them with the program's "automated" spell-checker, her agency should not have to perform an AIA for that spell-checker. Alternatively, an overly-narrow definition could undermine efforts to include high profile systems like those deciding where students go to school or how housing opportunities are allocated.<sup>31</sup>

It is also essential that "systems" are defined in terms that are broader than just their software — AIAs should address human and social factors, the histories of bias and discrimination in the context of use, and any input and training data.<sup>32</sup> Bias in automated decision systems can arise as much from human choices on how to design or train the system as it can from human errors in judgment when interpreting or acting on the outputs.<sup>33</sup> Evaluating a risk assessment tool, for instance, is not just a matter of understanding the math behind an algorithm; we must understand how judges, police officers, and other decision-makers influence its inputs and interpret its outputs.<sup>34</sup>

A reasonable education agency's definition, for example, should include an automated decision system such as the Educational Value-Added Assessment System, used by many jurisdictions for automated teacher evaluations.<sup>35</sup> The text of that agency's definition might include something like the "systems, tools, or statistical models used to measure or evaluate an individual teacher's performance or effectiveness in the classroom." In a criminal justice agency, similar wording might yield a definition that includes "systems, tools, or statistical models used to measure or evaluate an individual criminal defendant's risk of reoffending."

- 31 Atila Abdulkadiroğlu, Yeon-Koo Che, and Yosuke Yasuda, "Expanding" choice" in school choice," *American Economic Journal: Microeconomics* 7, no. 1 (2015): 1-42; Neil Thakral, "The Public-Housing Allocation Problem," Technical report, *Harvard University*, 2016. In the UK, a review of "governmental analytical models" focused on models that are used to inform agency decisions. The review, which went on to inform the UK Government's "Aqua Book" on guidance for producing quality analysis in government, offers one possible method for defining automated decision system ("Review of quality assurance of Government analytical models: final report," *HM Treasury, UK*, March 2013, <https://www.gov.uk/government/publications/review-of-quality-assurance-of-government-models>).
- 32 Aaron Reike, Miranda Bogen and David G. Robinson, *Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods*, (Upturn and Omidyar Network, 2018), <https://www.omidyar.com/insights/public-scrutiny-automated-decisions-early-lessons-and-emerging-methods>; April Glaser, "Who Trained Your A.I.," *Slate*, Oct. 24, 2017, [http://www.slate.com/articles/technology/technology/2017/10/what\\_happens\\_when\\_the\\_data\\_used\\_to\\_train\\_a\\_i\\_is\\_biased\\_and\\_old.html](http://www.slate.com/articles/technology/technology/2017/10/what_happens_when_the_data_used_to_train_a_i_is_biased_and_old.html).
- 33 Batya Friedman and Helen Nissenbaum, "Bias in computer systems," *ACM Transactions on Information Systems (TOIS)* 14, no. 3 (1996): 330-347. <https://www.nyu.edu/projects/nissenbaum/papers/biasincomputers.pdf>.
- 34 Steven L. Chanenson and Jordan M. Hyatt, "The Use of Risk Assessment at Sentencing: Implications for Research and Policy," *Villanova Law/Public Policy Research Paper* No. 2017-1040 (2017), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2961288](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2961288).
- 35 Moser, "A Controversial Teacher-Evaluation Method"; Cameron Langford, "Houston Schools Must Face Teacher Evaluation Lawsuit," *Courthouse News Service*, May 8, 2017, <https://www.courthousenews.com/houston-schools-must-face-teacher-evaluation-lawsuit/>.

### DEFINITIONS AND THE EU GENERAL DATA PROTECTION REGULATION

A definition of “automated decision system” that focuses on individual profiling has a precedent. In the European Union’s General Data Protection Regulation, automated profiling is defined as “any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.”<sup>36</sup>

The GDPR language may be a good starting point for some agencies, but will require some shaping to match the appropriate contexts. In other contexts it may not be sufficient. Some predictive policing tools, for example, do not necessarily “profile individuals,” and instead focus on locations, using statistics to try to understand and predict crime trends across geographical areas, with the potential for disparate impact. A definition might then have to account for “any systems, tools, or algorithms that attempt to predict crime trends and recommend the allocation of policing resources” in non-individualized terms. In general, any definition should certainly cover systems that might have a disparate impact on vulnerable communities and to pay careful attention to how broad terms, like “automated processing,” are specified in practice.

### **B. PUBLIC NOTICE OF EXISTING AND PROPOSED AUTOMATED DECISION SYSTEMS: ALERT COMMUNITIES ABOUT THE SYSTEMS THAT MAY AFFECT THEIR LIVES**

A fundamental aspect of government accountability and due process is notice of how our rights may be affected by government agencies and actors.<sup>37</sup> When automated systems play a significant role in government decisions, the public should be given notice. Substantive public engagement requires access to accurate and timely information. Thus, the second component of an Algorithmic Impact Assessment would require each agency to publicly disclose proposed and existing automated decision systems, including their purpose, reach, internal use policies, and potential impacts on communities or individuals.

This requirement by itself would go a long way towards shedding light on which technologies are being deployed and where accountability research and community advocacy should be focused. In response to concerns over the secretive and often unchecked use of new surveillance technology by local law enforcement, the City of Seattle adopted Ordinance 123576, which requires the public disclosure of city surveillance systems.<sup>38</sup> Even though city agencies have yet to start “backfilling”

36 “Definitions,” Art. 4, Regulation (EU) 2016/679, of the European Parliament and 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.

37 Citron, “Technological due process.”

38 Seattle’s “Surveillance Ordinance” requires agencies that acquire surveillance tools to publicly disclose those tools to the public, and create reports on the proposed use of those tools (Seattle, Washington, Surveillance Ordinance 123576, <http://seattle.legistar.com/ViewReport.ashx?M=R&N=Text&ID=393&ID=2849012&GUID=5B7D2F80-A918-4931-9E2E-88E27478A89E&Title=Legislation+Text>).

## II. THE CONTENT OF AN ALGORITHMIC IMPACT ASSESSMENT

Surveillance Impact Reports on technology the city already uses,<sup>39</sup> the city has begun to make the list of technologies publicly available in a simple, accessible format, allowing the public to raise informed questions.<sup>40</sup> The value of those simple disclosures alone can be a major asset for researchers and the public. Similar provisions are already part of laws in the U.S., such as the Privacy Act of 1974, and have been proposed in emerging local ordinances such as one in Santa Clara County and another in Oakland that are focused on privacy.<sup>41</sup>

AIA disclosures would also help governments proactively avoid political turmoil and backlash involving systems that the public may ultimately find untrustworthy or that may cause direct or indirect harm. For example, after investigative reporting revealed that the data analysis company Palantir had secretly partnered with the New Orleans Police Department on a predictive policing system that potentially reinforced racial and other biases,<sup>42</sup> the New Orleans Mayor's Office decided to allow the city's contract with Palantir to expire.<sup>43</sup> Had the New Orleans police department engaged in an AIA process before deployment, the system would have been subject to more rigorous review and possibly rejected outright, and many of the problems and objections might have been addressed without eroding public trust or possibly harming marginalized communities.<sup>44</sup>

### CHALLENGE: TRADE SECRECY

Public agencies will need to commit to accountability in both their internal technology development plans and their vendor and procurement relationships. For example, **the disclosure of automated decision systems and meaningful information about those systems will not be feasible if essential information is shielded from review by blanket claims of trade secrecy.**<sup>45</sup> While there are certainly some core aspects of systems that have competitive commercial value, it is unlikely that these extend to information such as the existence of the system, the purpose for which it was acquired, or the results of the agency's internal impact assessment.

Nor should trade secret claims stand as an obstacle to ensuring meaningful external research on such systems. AIAs provide an opportunity for agencies to raise any questions or concerns about trade secret claims in the pre-acquisition period, before entering into any contractual obligations. If a vendor objects to meaningful external review, this would signal a conflict between that vendor's system and public accountability. Such scenarios may require that

39 As of March 15, 2018.

40 "Surveillance Technologies," Seattle Information Technology, accessed March 16, 2018, <http://www.seattle.gov/tech/initiatives/privacy/surveillance-technologies>.

41 Ozer, "Santa Clara County Passes Landmark Law to Shut Down Secret Surveillance"; Darwin BondGraham, "Oakland Privacy Commission Approves Surveillance Transparency and Oversight Law," *East Bay Express*, Jan. 6, 2017, <https://www.eastbayexpress.com/SevenDays/archives/2017/01/06/oakland-privacy-commission-approves-surveillance-transparency-and-oversight-law>.

42 Winston, "Palantir."

43 Jonathan Bullington and Emily Lane, "New Orleans ends its relationship with tech firm Palantir, Landrieu's office says," *NOLA.com | The Times-Picayune*, March 14, 2018, [http://www.nola.com/crime/index.ssf/2018/03/palantir\\_new\\_orleans\\_gang\\_case.html](http://www.nola.com/crime/index.ssf/2018/03/palantir_new_orleans_gang_case.html).

44 Jessica Saunders, Priscillia Hunt, and John S. Hollywood, "Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot," *Journal of Experimental Criminology* 12:3 (2016), 347-371, <https://link.springer.com/article/10.1007/s11292-016-9272-0>.

45 Wexler, "Life, Liberty, and Trade Secrets"; Ram, "Innovating Criminal Justice".



## II. THE CONTENT OF AN ALGORITHMIC IMPACT ASSESSMENT

agencies ask potential vendors to waive restrictions on information necessary for external research and review.<sup>46</sup> At minimum, vendors should be contractually required by agencies to waive any proprietary or trade secrecy interest in information related to accountability, such as those surrounding testing, validation, and/or verification of system performance and disparate impact.<sup>47</sup> This also encourages a competitive landscape among government technology vendors to meet the accountability requirements of AIAs if they want to do business with public agencies.

### C. INTERNAL AGENCY SELF-ASSESSMENTS: INCREASE THE CAPACITY OF PUBLIC AGENCIES TO ASSESS FAIRNESS, JUSTICE, DUE PROCESS, AND DISPARATE IMPACT

Algorithmic Impact Assessments increase the internal capacity of public agencies to better understand and explicate potential impacts before systems are implemented.<sup>48</sup> Agencies must be experts on their own automated decision systems if they are to ensure public trust. This is why agencies' AIAs must include an evaluation of how a system might impact different communities and a plan for how agencies will address any issues, should they arise.

Ideally, government agencies should pre-identify issues and potential harms that will be evaluated in the self-assessment.<sup>49</sup> By standardizing the process, agencies can ensure the evaluation is comprehensive and comparable. The evaluation should be detailed so that outside researchers and experts can adequately scrutinize the system and its potential impact, and provide a non-technical summary for the general public. This dual explanation is used in other types of impact assessment frameworks and encourages robust public engagement.<sup>50</sup>

In their self-assessments, agencies should identify potential impacts on the public and then proactively engage affected communities to ensure that a system meets a given community's goals. The assessment should articulate why, in light of these goals, the system will have a net

46 David S. Levine, "The People's Trade Secrets," *18 Mich. Telecomm. & Tech. L. Rev.* 61 (2011), <https://repository.law.umich.edu/mttlr/vol18/iss1/2/>.

47 Jan Whittington, Ryan Calo, Mike Simon, and Jesse Woo, "Push, Pull, and Spill: A Transdisciplinary Case Study In Municipal Open Government," *30 Berkeley Tech. L.J.* 1967 (2015), <https://scholarship.law.berkeley.edu/btlj/vol30/iss2/2/>.

48 New York City Council, Hearing Testimony, Oct. 16, 2017, <http://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0>; The Federal Trade Commission. "Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues." January 2016. <https://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report>; David McCabe, "Lawmakers Are Trying to Understand How Tech Giants' Algorithms Work," *Axios*, Nov. 29, 2017, <https://www.axios.com/lawmakers-are-trying-to-understand-how-tech-giants-algorithms-work-1513307255-b4109efc-9566-4e69-8922-f37d9e829f1f.html>. ("We don't, I don't think, as a committee really know how to get the socks on the octopus, so to speak, here because it's complicated," said California Democrat Rep. Anna Eshoo, regarding increasing pressure for policymakers to understand big data and algorithms used by technology companies).

49 For example, in 2014, Former Attorney General Eric Holder urged the Sentencing Commission to "study the use of a data-driven analysis in front-end sentencing - and to issue policy recommendations based on this careful, independent analysis." Eric Holder, "Speech at the National Association of Criminal Defense Lawyers 57th Annual Meeting and 13th State Criminal Justice Network Conference" (Philadelphia, PA, Aug. 1, 2014), *Department of Justice*, <https://www.justice.gov/opa/speech/attorney-general-eric-holder-speaks-national-association-criminal-defense-lawyers-57th>.

50 John Fry, Anne Maxwell, Sarah Apere, Paddy McAweeney, Luke McSharry, and Ainhoa González, "Non-Technical Summaries-Due Care and Attention," *In 34th IAIA Annual Conference*, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.567.8444&rep=rep1&type=pdf>.

## II. THE CONTENT OF AN ALGORITHMIC IMPACT ASSESSMENT

positive impact on those communities.<sup>51</sup> Fulfilling this requirement of the AIA process would require an agency to engage those communities early on, even before the formal notice and comment process.

Agencies could also use the AIA as an opportunity to lay out any other procedures that will help secure public trust in such systems. If appropriate, the agency might want to identify how individuals can appeal decisions involving automated decision systems, to make clear what appeals processes might cover a given system's decision, or to share its mitigation strategy should the system behave in an unexpected and harmful way.<sup>52</sup> If a harm, an undesirable outcome, or an error is identified, the agency should explain how it intends to correct or remedy the issue.

This self-assessment process is also an opportunity for agencies to develop expertise when commissioning and purchasing automated decision systems, and for vendors to foster public trust in their systems. Agencies will be better able to assess the risks and benefits associated with different types of systems, and work with vendors and researchers to conduct and share relevant testing and research on their automated decision system, including but not limited to testing for any potential biases that could adversely impact an individual or group. Indeed, researchers are already developing resources and materials that agencies can use to ask appropriate questions of their own systems.<sup>53</sup> As noted above, if some vendors raise trade secrecy or confidentiality concerns, those can be addressed in the AIA, but responsibility for accountability ultimately falls upon the public agency.

The benefits of self assessments to public agencies go beyond algorithmic accountability: it encourages agencies to better manage their own technical systems and become leaders in the responsible integration of increasingly complex computational systems in governance.

### OPPORTUNITY: BENEFIT TO VENDORS

AIA's would also benefit vendors that prioritize fairness, accountability, and transparency in their offerings. Companies that are best equipped to help agencies and researchers study their systems would have a competitive advantage over others. Cooperation would also help improve public trust, especially at a time when skepticism of the societal benefits of tech companies is on the rise.<sup>54</sup> These new incentives can encourage a race to the top of the accountability spectrum among vendors.

51 Saunders, et al., "Predictions put into practice."

52 Danielle Keats Citron, "Big Data Should Be Regulated by 'Technological Due Process,'" *N.Y. Times*, July 29, 2016, <https://www.nytimes.com/roomfordebate/2014/08/06/is-big-data-spreading-inequality/big-data-should-be-regulated-by-technological-due-process>.; Citron, "Technological Due Process."; Citron & Pasquale, *The Scored Society*.; Crawford and Schultz. "Big data and due process."

53 Diakopolous, et al., "Principles for Accountable Algorithms and a Social Impact Statement for Algorithms," *FATML*, accessed March 16, 2018, <https://www.fatml.org/resources/principles-for-accountable-algorithms>.

54 Erin Griffith, "The Other Tech Bubble," *Wired*, Dec. 16, 2017, <https://www.wired.com/story/the-other-tech-bubble/>.



### OPPORTUNITY: AIAs AND PUBLIC RECORDS REQUESTS

Increasing agency expertise through AIAs will also help promote transparency and accountability in public records requests. Today, when agencies receive open records requests for information about algorithmic systems, there is often a mismatch between how the outside requestor thinks agencies use and classify these technologies and the reality.<sup>55</sup> As a result, requests may take a scattershot approach, cramming overly broad technical terms into numerous requests in the hopes that one or more hit the mark. This can make it difficult for records officers responding in good faith to understand the requests, let alone provide the answers the public needs.

Even open records experts who are willing to reasonably narrow their requests may be unable to do so because of the lack of any “roadmap” showing which systems a given agency is planning, procuring, or deploying. For example, in a project at the University of Maryland, faculty and students working in a media law class filed numerous general public records requests for information regarding criminal risk assessment algorithm usage in all fifty states.<sup>56</sup> The responses they received varied significantly, making it difficult to aggregate data and compare usage across jurisdictions. It also revealed a lack of general knowledge about the systems among the agencies, leading to situations where the students had to explain what ‘criminal justice algorithms’ were to the public servants in charge of providing the records on their use. Accountability processes such as the AIA would help correct this mismatch on both sides of the equation.

Researchers, journalists, legal organizations, and concerned members of the public could use AIAs to reasonably target their requests to systems that were enumerated and described, saving public records staff significant time and resources. Agency staff would also gain a better understanding of their own systems and records and could then help requestors understand which documents and public records are potentially available. This alignment would increase efficiency, lower the agency burden of processing requests, and increase public confidence. And of course, some basic requests will be preempted by the AIA’s disclosure requirement, saving researchers and the agencies the burden of engaging in the public records request process.

55 Katherine Fink, “Opening the government’s black boxes: freedom of information and algorithmic accountability, Information,” *Communication & Society* (2017), <https://www.tandfonline.com/doi/pdf/10.1080/1369118X.2017.1330418>.

56 Nicholas Diakopoulos, “We need to know the algorithms the government uses to make important decisions about us,” *The Conversation*, May 23, 2016, <https://theconversation.com/we-need-to-know-the-algorithms-the-government-uses-to-make-important-decisions-about-us-57869>.

### CHALLENGE: CONSIDERING BOTH ALLOCATIVE AND REPRESENTATIONAL HARMS

An anticipated challenge for governments performing Algorithmic Impact Assessments is the assessment of potential cultural and social harms. This challenge exists in other impact assessment processes because it requires the agency to make assumptions or predictions about cultural or social factors that vary enormously within and between communities and geographic areas. This practice often results in findings only reflecting potential impacts on a dominant culture and omitting or misinterpreting the impacts on marginalized communities and individuals. For instance, residents of a historically Black neighborhood shaped by Jim Crow segregation in Corpus Christi, Texas, reached a multi-million dollar settlement against the Federal Highway Administration because the Environmental Impact Assessment failed to anticipate that the highway construction plans further segregated this neighborhood.<sup>57</sup> Avoiding these sorts of harms is a key goal of the AIA notice and comment process.

The existing literature on bias in algorithmic systems has tended to rely heavily on what could be called “harms of allocation,” in which some groups are denied access to valuable resources and opportunities.<sup>58</sup> Of course, addressing allocative harms is crucial. But agencies should also consider harms of representation – the way a system may unintentionally underscore or reinforce the subordination of some social and cultural groups. For example, researchers classify Google’s photo platform’s automatic labeling of images of black people as “gorillas” as a representational harm,<sup>59</sup> and the denial of mortgages to people who live within a particular zip code as an allocative harm.<sup>60</sup> Automated decision systems used in the public sector are susceptible to both kinds of harm because they can be embedded with demographic data that serve as proxies for particular groups or reinforce past harms that can have economic or identity-based impacts.

### D. MEANINGFUL ACCESS: ALLOW RESEARCHERS AND AUDITORS TO REVIEW SYSTEMS ONCE THEY ARE DEPLOYED

**Algorithmic Impact Assessments should provide a comprehensive plan for giving external researchers and auditors meaningful, ongoing access to examine specific systems, to gain a fuller account of their workings, and to engage the public and affected communities in the process.** This plan should give experts rapid access to a system once it is deployed (e.g. within six months). However, in situations where internal agency assessments are insufficient or where particular risks or harms have gone

57 Lawyers’ Committee for Civil Rights Under Law, 2015, “Historic Agreement Resolves Environmental Justice Complaint In Corpus Christi, Texas,” <https://lawyerscommittee.org/press-release/historic-agreement-resolves-environmental-justice-complaint-in-corpus-christi-texas/>.

58 See the discussion of allocative and representational harms in: Solon Barocas, Kate Crawford, Aaron Shapiro and Hanna Wallach, “The Problem with Bias: From Allocative to Representational Harms in Machine Learning”, *SIGCIS* conference paper, October 2017; See also: Kate Crawford, “The Trouble with Bias”, *NIPS conference keynote*, December 2017, [https://www.youtube.com/watch?v=fMym\\_BKWQzk](https://www.youtube.com/watch?v=fMym_BKWQzk).

59 Tom Simonite, “When it comes to gorillas, Google photos remains blind,” *Wired*, January 11, 2018, <https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind>.

60 Kenneth R. Harney, “Zip code ‘redlining’: a sweeping view of risk,” *Washington Post*, February 2, 2008, <http://www.washingtonpost.com/wp-dyn/content/article/2008/02/01/AR2008020101680.html>.

## II. THE CONTENT OF AN ALGORITHMIC IMPACT ASSESSMENT

unaddressed, external researchers and auditors could raise the need for pre-deployment review in the comment period. While certain individuals and communities may wish to examine the systems themselves, this cannot be relied upon: it would be unreasonable to assume that everyone has the time, knowledge, and resources for such testing and auditing.<sup>61</sup> Automated decision systems can be incredibly complex, and issues like bias and systematic errors may not be easily determined through the review of systems on an individual, case-by-case basis.<sup>62</sup> A plan to grant meaningful access to qualified researchers would allow individuals and communities to call upon the trusted external experts best suited to examine and monitor a system to assess whether there are issues that might harm the public interest.<sup>63</sup>

To do this well, it is important to recognize that the appropriate type and level of access may vary from agency to agency, from system to system, and from community to community. The risks and harms at issue in different systems may demand different types of assessment and auditing using different methods and disciplines. While the right to an explanation concerning a specific automated decision could prove useful in some situations, many systems may require a group-level or community-wide analysis. For example, an explanation for a single “stop and frisk” incident would not reveal the greater discriminatory pattern that the policy created in New York City, where over 80% of those stopped were Black or Latino men.<sup>64</sup>

Many systems may only require analysis based on inputs, outputs, and simple information about the algorithms used without needing access to the underlying source code.<sup>65</sup> We expect that for many systems, agencies would have to provide training data or a record of past decisions to researchers. We believe that the best way for agencies to develop an appropriate research access process initially would be to work with community stakeholders and interdisciplinary researchers through the notice and comment process. Importantly, given changing technologies, the developing research field around accountability, and the shifting social and political contexts within which systems are deployed, access to a system will almost certainly need to be ongoing, and take the form of monitoring over time.<sup>66</sup>

As an individual agency works with researchers and community members to design its research access provisions, there are a number of elements that should be in place. Research and auditing performed on these systems should be accountable to the public, and should include a public log of which researchers and experts are provided access, and on what basis. Agencies should ensure that affected communities are able to suggest researchers that they feel represent their interests, and should work with researchers to

61 Mike Ananny and Kate Crawford, “Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability,” *New Media & Society* (2016)

62 Diakopoulos, “Algorithmic Accountability” (discussing variable sources of error and correspondingly myriad approaches to algorithmic accountability).

63 Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort, “Auditing algorithms: Research methods for detecting discrimination on internet platforms,” *Data and discrimination: converting critical concerns into productive inquiry* (2014): 1-23.; Devin G. Pope and Justin R. Sydnor. “Implementing anti-discrimination policies in statistical profiling models.” *American Economic Journal: Economic Policy* 3, no. 3 (2011): 206-31.

64 “Stop-and-Frisk Data,” NYCLU, accessed March 16, 2018, <https://www.nyclu.org/en/stop-and-frisk-data>.

65 Kristian Lum and William Isaac, “To predict and serve?,” *Significance* 13, no. 5 (2016): 14-19. <http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2016.00960.x/full>.

66 *Conference on Fairness, Accountability, and Transparency*, <https://fatconference.org>.

## II. THE CONTENT OF AN ALGORITHMIC IMPACT ASSESSMENT

ensure that these communities have a voice in formulating the questions that are asked and addressed by research and auditing. Importantly, to ensure public accountability and a thriving research field, research findings and conclusions should be published openly (even if after an embargo period), and be held to standards of scrutiny and peer review within the appropriate research domains.

Ongoing auditing and research access would allow agencies, researchers, and affected communities to work together to develop their approaches to testing and interrogating these systems. This is especially important given that the research about algorithmic accountability is young and technological development proceeds rapidly. We do not yet know what future tools, techniques, and perspectives might best keep systems accountable. External experts from a wide variety of disciplines will need the flexibility to adapt to new methods of accountability as new forms of automated decision making emerge.<sup>67</sup>

### CHALLENGE: FUNDING AND RESOURCES

Of course, there is also a real danger that relying on external auditing will become an unfunded tax on researchers and the affected communities they engage with, who might be expected to take responsibility for testing and monitoring automated decision systems without resources or compensation. Alternatively, if in-house auditors are relied on, they could become captured by the incentives of their clients or face conflict-of-interest issues. However, there are approaches that legislation could adopt to address this. An AIA framework could fund an independent, government-wide oversight body, like an inspector general's office, to support the research, access, and community engagement.<sup>68</sup> Community institutional review boards could be supported to help steer and review research proposals.<sup>69</sup> Funding could be set aside for the compensation of external auditors. Fortunately, there are many options that jurisdictions could consider for their own needs. A growing community of computer scientists, journalists, social scientists, and engaged community advocates have already proven there is an appetite for research into public automated systems. This work should continue to be strongly supported by funding bodies and research agencies.

67 AI Now 2016 Symposium, July 7, 2016, <https://ainowinstitute.org/events/2016-symposium.html>; AI Now 2017 Symposium, July 10, 2017, <https://ainowinstitute.org/events/2017-symposium.html>.

68 Executive Office of the President, *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*, May 2016, [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016\\_0504\\_data\\_discrimination.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf); U.S. Department of Justice, Office of the Inspector General (2018), <https://oig.justice.gov>.

69 Phoebe Friesen, Lisa Kearns, Barbara K. Redman and Arthur L. Caplan, "Extending Ethical Strides: From Tribal IRBs to the Bronx Community Research Review Board," *The American Journal of Bioethics* (2017), 17:11, W5-W8, <https://www.tandfonline.com/doi/abs/10.1080/15265161.2017.1378755>.

### **III. CONCLUSION**

As more governments adopt automated decision systems, public agencies will need a way to address the accompanying risks to fairness, justice, and due process, and to include affected communities in the conversation. Algorithmic Impact Assessments offer agencies a framework for understanding the automated decision systems they procure, and give the public more insight into the workings of automated decision systems in order to keep them accountable. Through public notice of system adoption, agency self-assessment, a plan for meaningful access for researchers and experts, and due process mechanisms, AIAs will help to ensure that governments are ready to face the risks presented by automated decision systems.

# ACKNOWLEDGEMENTS

Thank you to **Amanda Levendowski** and **Rashida Richardson** for help drafting and editing this report.

Thank you to **Stephanie Ballard, Chris Bavitz, Esha Bhandari, Hannah Bloch-Wehba, Rachel Brooke, Ryan Calo, Craig Campbell, Corinne Cath, Danielle Citron, Kade Crockford, Cassie Deskus, Ed Felten, Rachel Goodman, Samantha Grassle, Chuck Howell, Shankar Narayan, Nicole Ozer, Frank Pasquale, Julia Powles, Eric Sears, Andrew Selbst, Jake Snow, Vincent Southerland, Tony Thompson, Michael Veale, Rebecca White**, and attendees of Princeton CITP's AI and Ethics Conference for their helpful comments on the AIA framework.