Automated Decision-making Systems Workgroup

WaTech
Washington Technology Solutions
Washington's Consolidated Technology Services Agency

Office of the Chief Information Officer

Automated Decision-Making Systems Workgroup Report

December 2021

Table of Contents

Proviso 2
Background 4
Purpose 4
Membership 4
Proviso Directions 6
Meetings 6
System selection 6
Methods of ADS Workgroup 6
System Review 8
Research Findings Summary 9
Guiding Principles 10
Recommendations 10
Challenges for policy change considerations: Definition of an ADS system 13
Effect and Evaluation of Recommendations on WA ONE System 14
Relevant Proviso Details 16
Definition of automated decision system 16
Authorizing statute 16
Contact 16
Appendix A: ADS Research Findings 17
Appendix B: Guiding Principles for Governmental Entities Using Automated Decision-Making Systems 21
Sample Prioritization Matrix 24
Appendix C: Complete Questions and Answers regarding WA ONE System 25
Appendix D: ADS legislation, regulation, and policy in other jurisdictions 29
In the United States 30
Around the world 34
Frameworks for determining whether uses of ADS or AI-enabled profiling should be prohibited 40
Glossary 42
Acknowledgments 43

Proviso

(14) $12,000 of the general fund—state appropriation for fiscal year 2022 is provided solely for the office of the chief information officer who must convene a work group to examine how automated decision-making systems can best be reviewed before adoption and while in operation and be periodically audited to ensure that such systems are fair, transparent, accountable and do not improperly advantage or disadvantage Washington residents.

a) The work group must be composed of:

   i. A representative of the department of children, youth, and families;

   ii. A representative of the department of corrections;

   iii. A representative of the department of social and health services;

   iv. A representative of the department of enterprise services;

   v. At least two representatives from universities or research institutions who are experts in the design and effect of an algorithmic system; and

   vi. At least five representatives from advocacy organizations that represent communities that are disproportionately vulnerable to being harmed by algorithmic bias, including but not limited to, African American, Hispanic American, Native American, and Asian American communities, religious minorities, people with disabilities, and other vulnerable communities.

b) The purpose of the work group is to develop recommendations for changes in state law and policy regarding the development, procurement, and use of automated decision systems by public agencies.

   The work group must examine:

   i. When state agency use of automated decision-making systems should be prohibited;

   ii. When state agency use of artificial intelligence-enabled profiling systems should be prohibited;

   iii. Changes in the procurement of automated decision systems, including when the procurement must receive prior approval by the office of chief information officer;

   iv. How to review, identify, and audit systems to ensure that the system prior to procurement and after placed into service does not discriminate against an individual, or treat an individual less favorably than another, in whole or in part, on the basis of one or more factors enumerated in RCW 49.60.010;

   v. How to provide public notice when an automated decision system is in use and how to appeal such decisions;

   vi. How automated decision system data should be stored and whether such data should be shared outside the system; and

vii. Other issues determined by the office of chief information officer or the department of enterprise services that are necessary to govern state agency procurement and use of automated decision systems.

c) To demonstrate the impacts of its recommendations, the work group must select one of following automated decision-making systems and describe how their implementation would affect the procurement of a new system and the use the existing system:

   i. The department of children, youth, and families system used to determine risk in the family child welfare system;

   ii. The department of corrections system used to determine risk for purposes of evaluating early release and/or sentencing; or

   iii. The department of social and health services system used for hospital admissions.

d) The work group shall meet at least four times, or more frequently to accomplish its work. The office of the chief information officer must lead the work group. Each of the state agencies identified in (a) of this subsection must provide staff support to the work group and its activities.

e) The work group must submit a report to the fiscal committees of the legislature and the governor no later than December 1, 2021.

f) For purposes of this subsection, "automated decision system" or "system" means any algorithm, including one incorporating machine learning or other artificial intelligence techniques, that uses data-based analysis or calculations to make or support government decisions, judgments, or conclusions that cause a Washington resident to be treated differently than another Washington resident in the nature or amount of governmental interaction with that individual including, without limitation, benefits, protections, required payments, penalties, regulations, timing, application, or process requirements.

Link to proviso: https://lawfilesext.leg.wa.gov/biennium/2021-22/Pdf/Bills/Session%20Laws/Senate/5092-S.SL.pdf#page=116

Background

In the 2021-23 biennial operating budget, the Washington State Legislature provided one-time funding of $12,000 to the Office of the Chief Information Officer (OCIO) to convene a workgroup and produce a report related to the adoption and use of automated decision-making systems (ADS) in the state. Due to the nature of automated decision systems and the increasing adoption of emerging technology in government, it is important that these systems are deployed in a fair, transparent and accountable manner.

Purpose

The purpose of the ADS Workgroup is to develop recommendations for changes in state law and policy regarding the development, procurement, and use of automated decision systems by public agencies. The ADS Workgroup examined how automated decision-making systems can best be reviewed before adoption, and while in operation, and be periodically audited to ensure that such systems are fair, transparent, and accountable and do not improperly advantage or disadvantage Washington residents. This report serves as the deliverable required by the budget proviso.

Membership

The budget proviso required the following membership make-up in the workgroup.

- A representative of the OCIO, who chairs the workgroup.
- A representative of the Department of Children, Youth, and Families (DCYF).
- A representative of the Department of Corrections (DOC).
- A representative of the Department of Social and Health Services (DSHS).
- A representative of the Department of Enterprise Services (DES).
- At least two (2) representatives from universities or research institutions who are experts in the design and effect of an algorithmic system.
- At least five (5) representatives from advocacy organizations that represent communities that are disproportionately vulnerable to being harmed by algorithmic bias, including, but not limited to, African American, Hispanic American, Native American, and Asian American communities, religious minorities, people with disabilities, and other vulnerable communities.

Based on these requirements the ADS Workgroup Membership was chaired by State Chief Privacy Officer Katy Ruckle, J.D., from the OCIO, and Dr. David D. Luxton, PhD., formerly Director of Research and Data Analytics at Department of Corrections and Director of Counseling & Wellness Programs, the Washington Department of Veterans Affairs. Workgroup members were from state agencies and advocacy organizations. Please see table below for the full member list:

	Last Name	First Name	Organization	Membership Status
1	Pincus	Jon	A Change Is Coming	Research Expert
2	Lee	Jennifer	ACLU - American Civil Liberties Union	Advocacy
3	Gonzalez	Eric	ACLU - American Civil Liberties Union	Advocacy
4	Block	Bill	ACLU - American Civil Liberties Union	Advocacy
5	Aguilar	Nancy	CHA - Commission on Hispanic Affairs	Advocacy
6	Auffray	Brianna	CAIR-WA - Council on American-Islamic Relations Washington	Advocacy
7	Krutsinger	Allison	DCFY - Department of Children Youth and Families	Required Agency
8	Mason	Aaron	DCFY - Department of Children Youth and Families	Required Agency
9	Ybarra	Vickie	DCFY - Department of Children Youth and Families	Required Agency
10	McGrew	Elena	DES - Department of Enterprise Services	Required Agency
11	Japhet	Robin	DES - Department of Enterprise Services	Required Agency
12	Fisher	Greg	DOC - Department of Corrections	Required Agency
13	Luxton	David	DOC - Department of Corrections (moved to DVA 9/1)	Required Agency
14	Adams	Gena	DOC - Department of Corrections	Required Agency
15	Bagdon-Cox	Courtney	DOC - Department of Corrections	Required Agency
16	Palma	Sergio	DSHS/ALTSA - Department of Social and Health Services/Aging and Long-Term Services Administration	Required Agency
17	Gogan	Jenise	DSHS/BHA - Department of Social and Health Services/Behavioral Health Administration	Required Agency
18	Mancuso	David	DSHS/RDA - Department of Social and Health Services/Research and Data Analysis	Required Agency
19	Henson	Crystal	DVA - Department of Veteran Affairs	Optional Agency
20	Allred	Robert	ESD - Employment Security Department	Optional Agency
21	Gordon	Elizabeth	Governor's Committee for Disability Issues and Employment	Advocacy
22	Chen	Christopher	HCA - Health Care Authority	Optional Agency
23	Ott	Cathie	HCA - Health Care Authority	Optional Agency
24	Del Villar	Ashley	La Resistencia and Mijente	Advocacy
25	Glenn	Kirsta	LNI - Labor and Industries	Optional Agency
26	Ruckle	Katy	OCIO - Office of the Chief Information Officer	Required Agency
27	Angel	Maria	University of Washington Law	Research Expert
28	Puckett	Derek	WaTech - Consolidated Technology Services	Required Agency

Proviso Directions

The 2021-2023 biennial budget proviso laid out clear deliverables of the ADS Workgroup, as well as processes and definitions. The Workgroup was primarily tasked with the following activities:

- Develop recommendations for changes in state law and policy regarding the development, procurement, and use of automated decision systems by public agencies.
- Assess the impact of those recommendations by evaluating an existing system in use by an agency as described within the proviso.
- Produce a final report to the Legislature and Governor by December 1, 2021.

Meetings

The budget proviso directed the ADS Workgroup to meet at least four times, or more frequently if necessary, to complete its work. Given the breadth and complexity of the work assigned, the Workgroup elected to meet every other week and held a total of 10 two-hour meetings. All meetings were recorded and open for public participation. Meeting agendas, slide decks, notes, and recordings of the meetings are posted on the ADS project website at https://watech.wa.gov/privacy/projects-and-initiatives.

System selection

To demonstrate the impact of its recommendations, the budget proviso required the ADS Workgroup to select a system in use by a state agency and describe how the implementation of any recommendations would affect:

(1) The procurement of a new system; and

(2) The use of an existing system.

For this work, the ADS Workgroup was required to select one of the following systems:

- The Department of Children, Youth, and Families (DCYF) system used to determine risk in the family child welfare system;
- The Department of Corrections (DOC) system used to determine risk for purposes of evaluating early release and/or sentencing; or
- The Department of Social and Health Services (DSHS) system used for hospital admissions.

Methods of ADS Workgroup

In the kick-off meeting, the ADS Workgroup heard presentations on the DCYF, DOC, and DSHS systems. Workgroup members then voted on which system to review. The Department of Corrections Washington Offender Needs Evaluation (WA ONE) system was selected. The WA ONE system identifies risk level classifications for all individuals under DOC jurisdiction - not just those housed at DOC facilities. It is used to make determinations concerning risk-based supervision. The assessment performed by the system is required by RCW 72.09.270.

Due to the wide variety of ADS Workgroup member backgrounds, the Workgroup also heard presentations by experts in this field who described and discussed the various aspects of automated decision-making systems, artificial intelligence technology, and procurement. These presenters and presentations included in order of appearance:

Santosh Putchala presented on July 29, 2021 on *Artificial Intelligence (AI) Primer and Algorithmic Bias*

Santosh Putchala is Director of Privacy at Kuma, LLC, a privacy, security, and identity management consultancy currently supporting the Washington state Office of Privacy and Data Protection. He is an industry leader in privacy and security, and is actively engaged in shaping the policy, strategy, implementation, of privacy and data protection. His work over the past 15 years spans privacy, security, data protection, compliance, regulatory guidance, standards development, certifications, and product management. Santosh regularly advises government, commercial and non-profit entities across various sectors and domains. He also holds degrees in engineering and law, and advanced degrees in cyber law, cybersecurity, and consumer privacy protection.

Elana McGrew presented on August 12, 2021 on an *Overview of Procurement in Washington.*

Elena McGrew is an Acting Statewide Enterprise Procurement Manager for the Contracts and Procurement Division (C&P) at the Department of Enterprise Services. Elena has been with C&P since 2015. Elena leads a team of Procurement Strategists, who consult on complex state acquisitions, and create tools and procedures to improve statewide procurement practices.

Ryan Calo, J.D., presented on August 12, 2021 on *How Policymakers Should or Shouldn't Use AI to Make Decisions*

Ryan Calo is the Lane Powell and D. Wayne Gittinger Professor at the University of Washington School of Law. He is a founding co-director of the interdisciplinary UW Tech Policy Lab and the UW Center for an Informed Public. Professor Calo holds adjunct appointments at the University of Washington Information School and the Paul G. Allen School of Computer Science and Engineering. He is an expert in law and technology, especially privacy, artificial intelligence, and robotics.

David D. Luxton, PhD., presented on August 26, 2021 on *Risks and Benefits of Automated Decision-Making Systems: A Preliminary Framework for Guiding Evaluation*

Dr. David D. Luxton is a clinical psychologist and research scientist who has authored over 100 academic articles in fields of artificial intelligence, ethics, and psychological health. He is also an Affiliate Associate Professor in Psychiatry and Behavior Sciences at the University of Washington and previously worked for the Washington State Department of Corrections. He currently serves as the Director of Counseling & Wellness Programs at the Department of Veteran Affairs.

Jon Pincus presented on September 23, 2021 on a *Discussion of what (if any) ADS uses should be prohibited?*

Jon Pincus is a researcher, entrepreneur, and strategist. His career includes several startups and almost a decade at Microsoft, first in Microsoft Research and then as General Manager for Strategy Development. His current project is The Nexus Today, a news site that uses anti-oppressive algorithms and design that highlights news and perspectives that are usually marginalized.

All of the informative presentations from these esteemed individuals are available on the project website at https://watech.wa.gov/privacy/projects-and-initiatives.

System Review

After much work and discussion regarding the review, the ADS Workgroup elected to review the DOC Washington Offender Needs Evaluation (WA ONE) system using questions drafted by the Americans Civil Liberties Union (ACLU). The questions and answers from DOC were then distributed to Workgroup members for review, which resulted in several more questions from Workgroup members and responses from the DOC. The Sept. 23, 2021, Workgroup meeting was dedicated to a review and in-depth discussion of the WA ONE system.

For reference, the WA ONE system classifies incarcerated adults and adults under community supervision for risk of recidivism and need areas. The risk and needs assessment first identifies need areas which affects program participation and prioritization. Since different programs are offered at different facilities, this can result in impacts to the place of incarceration. Second the assessment predicts potential recidivism and groups individuals into tiers of potential risk, which is the primary factor for determining frequency and types of community supervision contacts. Approximately 27,000 adults are affected by their WA ONE classification at any given time. DOC has been mandated by the Legislature to adopt a risk-assessment system recommended by the Washington State Institute for Public Policy (WSIPP). The content of the system and its testing, monitoring and transparency is the responsibility of the DOC.

Main themes and concerns about the Washington ONE system included:

- DOC's level of involvement in the design process of the system.
- Clarification regarding the metrics collected or used by the system.
- Scoring, sharing and use of risk scores, and permanence of those scores.
- Transparency of the system algorithm to the public, system owners, auditors, etc.
- Intended and unintended differential effects (i.e., gender-responsive versus gender distinctive tools, measuring racial bias).
- Updating algorithms in response to legislative and/or population changes (i.e., *Blake Decision*).

Concerns about the ADS review questions included:

- The underlying approach for selecting metrics and/or developing algorithms developed for an ADS.
- Studying bias as a best practice and the adequacy of bias testing.
- Phrasing of questions about "bias testing."
    - Feedback from the DOC indicated that rephrasing Question 9 in the assessment to include "algorithmic auditing" would not have had an impact on their response.
- Intended function and design of a system and potential sources or indicators of bias.

- Streamlining questions and reducing the time taken to complete a review of an ADS.

For a complete list of questions and answers, please see Appendix B to this report.

Research Findings Summary

- Governmental entities are increasingly using automated decision systems to automate or support both simple and complex decision-making processes.[1] These systems use algorithms, or a series of steps, to transform inputs into outputs that support or make decisions.
- Automated decision systems are a way to reduce costs, improve delivery of public services, and make decisions more efficient, reliable, and accurate. In some cases, use of automated decision systems may be mandated in legislation. However, a growing body of evidence indicates that automated decision systems can be discriminatory, inaccurate and lack transparency and accountability.[2]
- Washington state agencies currently deploy a large number and range of automated decision systems – some of which use simple rules-based algorithms and others that incorporate machine learning. It is important to note that the benefits and risks posed by use of these systems depend on the specific social, political, and institutional contexts in which they are deployed. Even simple rules-based algorithmic systems can pose significant risks if they affect many people or if they support high-stakes decisions. More complex machine-learning based systems can introduce additional risks such as the effects of being trained on biased or discriminatory data.
- In recent years, there has been a rapid evolution in the understanding of how algorithmically driven automated decision systems operate as well as the risks posed by their use. Some systems in use by governmental agencies today have not been audited for biases, and in many cases, were developed several years ago when techniques for identifying and addressing biases were not as advanced as they are today. An increasing number of jurisdictions are regulating automated decision systems (see ADS landscape section in Appendix D), understanding that while governmental entities may seek the benefits that these systems to deliver services, there are risks of using such systems that should be addressed to prevent harm.
- Washington state has introduced one of the country's first pieces of legislation to consider application and use of automated decision systems in government. Washington can be a leader in ensuring that government use of automated decision systems does not cause discrimination and other types of harm. (See Appendix A for full ADS research findings.)

[1] https://ainowinstitute.org/nycadschart.pdf

[2] https://scholarlycommons.law.emory.edu/cgi/viewcontent.cgi?article=1418&context=elj

Guiding Principles

The ADS Workgroup's recommendations follow from the Guiding Principles, derived from the expert testimony it heard, articles it reviewed and from legal and resource considerations. Overall, the ADS Workgroup accepts that the practical and policy considerations behind the procurement, implementation, and management of ADS systems are best made by keeping the Guiding Principles in mind. The Guiding Principles should be viewed as the goals and key considerations to be used by governmental entities procuring, implementing, and managing their ADS. Weighing how and when to translate goals into reality is the important work of government. The Guiding Principles address the following:

1. Prioritization.
2. Evaluation Whether to Adopt System.
3. Updated Assessments.
4. Periodic Implementation Testing.
5. Transparency of the Algorithm.
6. Audit Trails.
7. Training on Automation Bias.
8. Evaluation of Risks/Determination Whether to Proceed.
9. Review of Decisions by those Affected.
10. Weighing Advantages Against Known Bias or Inaccuracies.
11. Review of Current Systems and Processes/Necessary Action.

For a complete description of the Guiding Principles see Appendix B.

Recommendations

The budget proviso specifically requested recommendations related to the following questions:

- Are there changes needed regarding the development, procurement, and use of ADS by state agencies?
    - If yes, what types of changes regarding:
        - Development?
        - Procurement?
        - Use?
- How can ADS be reviewed before adoption?
- How can ADS be reviewed while in operation?
- How can systems be audited to ensure ADS is fair, transparent, and accountable?

- How can the state ensure ADS does not improperly advantage or disadvantage particular residents?
- Are there circumstances when state agency use of automated decision-making systems should be prohibited?
- Are there circumstances when state agency use of artificial intelligence-enabled profiling systems should be prohibited?

Below are the recommendations from the ADS Workgroup based on our discussions and review of the Washington ONE system.

Recommendation #1 Prioritization of Resources

The state should develop a prioritization framework (see example in Appendix B) for allocating resources to address existing and future ADS. While the prioritization framework is being developed, agencies should adopt and implement an interim prioritization framework so as not to delay work relating to ADS systems. The prioritization framework should be used in determining the level of resources to be devoted to meeting each of the following recommendations.

Recommendation #2 Procurement

As a part of the procurement process, assess new automated decision systems procured by the state. The assessment should include evaluation of the potential impacts of the automated decision-making on (1) the risk to rights and freedoms to an "identified or identifiable natural person," (2) the existence or risk of bias or inaccuracy in the results of the system and (3) whether the workings of the system are transparent to the public.

"Identified or identifiable natural person" means a person who can be readily identified, directly or indirectly. This definition applies in the other recommendations as well.

Recommendation #3 Evaluation of Existing Systems

Automated decision-making systems currently in use by the state that produce legal effects on identified or identifiable natural persons should be assessed if they are processing data on a large scale or have substantial effects on the rights or freedoms of natural persons. The assessment should include the existence or risk of bias or inaccuracy in the results and how transparent the system use, and impacts are to the public.

Recommendation #4 Transparency

Require transparency of use, procurement, and development of automated decision-making systems, including monitoring or testing for accuracy and bias, that produce legal effects on identified or identifiable natural persons.

Recommendation #5 Determination on Whether to Use System

The state should adopt a framework to evaluate state agency use of ADS technology or use of artificial intelligence-enabled profiling to determine whether or not its use should be prohibited.[3]

Recommendation #6 Ongoing Monitoring or Auditing

Ongoing monitoring or auditing should be performed on ADS systems that have legal effects on identified or identifiable natural person's to ensure they do not have differential effects on subpopulations that result over time; or discriminate against an individual, or treat an individual less favorably than another, in whole or in part, on the basis of one or more factors enumerated in RCW 49.60.010.

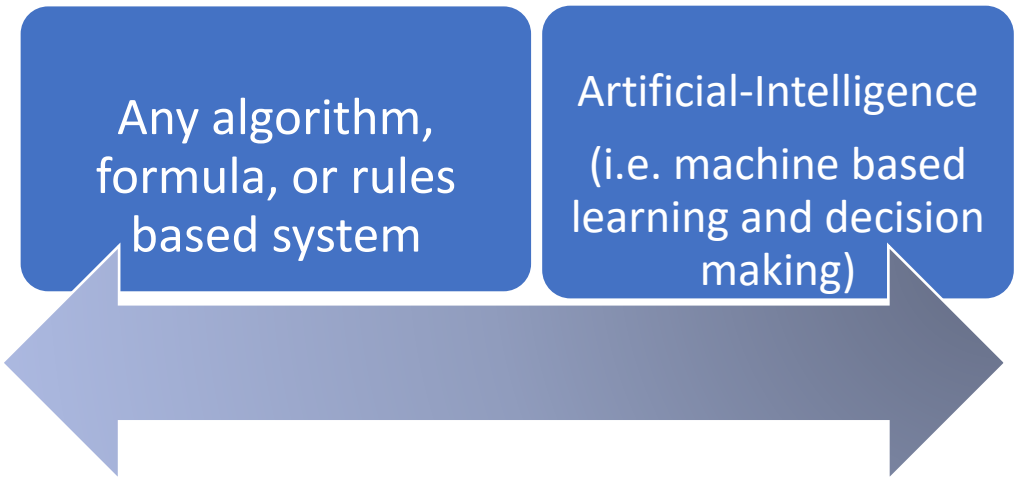
Recommendation #7 Training in Risk of Automation Bias

Require training of state employees who develop, procure, operate or use automatic decision-making systems as to risk of automation bias.

[3] The ADS Workgroup is not recommending any specific framework currently. Appendix D on the ADS Policy Landscape discusses frameworks that have been used elsewhere, including whether a technology endangers civil liberties, violates fundamental rights, or exacerbates racial injustice.

Challenges for policy change considerations: Definition of an ADS system

Even with the definition[4] provided within the budget proviso of what an "automated decision-making system" is there was not consensus within the Workgroup of what systems were meant to be within scope of the definition. The definition provided in the proviso was the source of many hours of discussion regarding what state agencies should consider to be in scope. Some Workgroup members expressed a desire to include any system in scope that uses any algorithm to make or assist in making decisions. This would include applications with built-in formulas such as Excel spreadsheets. The reasoning for this widely inclusive interpretation is the concern that any system has the potential to inflict harm and disparate impacts. This group would rely on the prioritization matrix (see Guiding Principles Appendix B) to ensure that the broad definition did not divert resources to low-risk, low-effect systems. Other Workgroup members were less sure that the scope and definition should be as broad and were concerned that interpreting the definition too broadly would bring too many inconsequential systems in scope and so would use a narrower definition. It can be understood on a spectrum with the broad interpretation on one-end which includes any system that calculates decisions to what is understood as actual "artificial-intelligence" or AI on the other end. AI can be understood as a branch of computer science that creates algorithms to classify, analyze, and draw predictions from data that may mimic human behavior based on the data sets used to train the AI.



Despite these definition and scope challenges, the Workgroup did agree that the state does not want systems that perpetrate patterns of bias or discrimination against Washington residents.

[4] See definition in Relevant Proviso Details.

Effect and Evaluation of Recommendations on WA ONE System

The ADS Workgroup proviso direction includes selecting an ADS from among the three agencies listed (DCYF, DOC, and DSHS) and describing how implementation of the Workgroup Recommendations would affect the procurement of a new system and the use of the existing system. The ADS Workgroup chose the Department of Corrections (DOC) Washington Offender Needs Evaluation (WA ONE) system, which classifies incarcerated adults and adults under community supervision for risk of recidivism and need areas.

The risk and needs assessment first identifies need areas which affects program participation and prioritization. Since different programs are offered at different facilities, this can result in impacts to place of incarceration. Second the assessment predicts potential recidivism and groups individuals into tiers of potential risk, which is the primary factor for determining frequency and types of community supervision contacts. Approximately 27,000 adults are affected by their WA ONE classification at any given time. The DOC has been mandated by the Legislature to adopt a risk-assessment system recommended by the Washington State Institute for Public Policy (WSIPP). The content of the system and its testing, monitoring and transparency is the responsibility of the DOC.

The ADS Workgroup Recommendations are based on the Guiding Principles. The following analysis focuses on the Recommendations and addresses them in numbered order.

1. Prioritization: The Recommendations call for establishment of first interim and then a final prioritization framework to determine the level of resources to be devoted to meeting the rest of the Recommendations. The WA ONE system would be a priority system under virtually any prioritization matrix. It affects a very large number of persons (all those incarcerated or under community supervision) and it affects significant rights and freedoms, determining both programming needs of persons while incarcerated and the level of reporting for community supervision.
    - Effect of Recommendations: Under the Recommendations, WA ONE would be a priority system with regard to allocation of resources to examine its procurement and operation.
2. Procurement: The recommendations state that during procurement, there should be evaluation of the risk of inaccuracy or bias in the results and whether the system is transparent to the public. The Workgroup received a simple assertion that in design WA ONE somehow went through “initial analysis of biases,” but any work done was performed by the vendor and has not been published, so whether or to what extent there was any rigorous testing is extremely uncertain. There was no involvement of the public in design of the system, only a post-design engagement to “answer questions” with the Family Council. The vendor has claimed the algorithm as proprietary and therefore not disclosable, so the system is not transparent to the public and has only limited transparency even to the agency responsible for implementing it. There was an assertion that a greater disclosure was made to WSIPP in connection with its currently pending study, but that pending study does not include any evaluation for bias.

- **Effect of Recommendations:** Under the Recommendations, there would have been a rigorous evaluation for inaccuracy and bias as part of the procurement process (either before procurement or during development of the instrument and before implementation), and according to the Guiding Principles, where, as here, the system is a high priority system and the algorithm is not available to the public, that evaluation would have been made by an independent third party.

3. **Evaluation of Existing Systems:** Based on the information available to the ADS Workgroup, implementation of the WA ONE system has not included adequate monitoring or testing for bias. There has been periodic testing for accuracy, which has found that the accuracy of prediction of recidivism is greater for WA ONE than for certain other systems tested (although the level of accuracy – an "Area Under the Curve" of .72 – involves considerable inaccuracy).[5] The WA ONE system has not yet been subject however to the types of bias analyses including subgroup bias analysis currently recommended and becoming standard practice for a system of this type.
   - **Effect of Recommendations:** The studies that have been made have focused on accuracy and cost. The limited review of bias to date has been inadequate and even then shows that potential disparities exist. Under the Recommendations this high-priority system should be subject to ongoing periodic evaluation for bias.
4. **Transparency:** The algorithm has been deemed by the vendor proprietary so there has been essentially no transparency in development. The procurement was by open bid, but without public comment or input. The implementation has been subject to several studies for accuracy by the system vendor and one meta-comparison of system studies by WSIPP (which did not independently evaluate accuracy), and WSIPP has apparently been given some access to the algorithm, but only to examine accuracy. Otherwise use and monitoring has not been transparent.
   - **Effect of Recommendations:** The Recommendations are for transparency in use, procurement and development and in monitoring or testing for accuracy or bias. In the case of WA ONE, that transparency has been limited. As recognized in the Guiding Principles, the appropriate compensation for the lack of transparency in a high-priority system is robust, independent, third-party monitoring and testing for accuracy and bias.
5. **Determination on whether to use System**: The existence of a risk evaluation system has been mandated by the Legislature.
   - **Effect of Evaluation:** The adoption of a system is consistent with the legislative mandate and is not optional.
6. **Ongoing Monitoring or Auditing:** As noted above, there is only one chart that addresses racial discrimination or differential effects, and it is incomplete in several

[5] Predicting Criminal Recidivism, Washington State Institute for Public Policy (February 2014).

critical respects. The 2021 WSIPP study was initially described to the Workgroup as involving a review of bias, but apparently will not do so.

- **Effect of Recommendations:** The Recommendations are for ongoing monitoring or auditing to ensure that WA ONE does not have differential effects on subpopulations that result over time; or discriminate against an individual, or treat an individual less favorably than another, in whole or in part, on the basis of one or more factors enumerated in RCW 49.60.010.

7. **Training in Risk of Automation Bias:** It is not clear whether DOC personnel involved in WA ONE are involved in training on automation bias.
    - **Effect of Recommendations:** Under the Recommendations the DOC would provide personnel involved in procurement and operation of the risk-assessment system with training on automation bias.

Relevant Proviso Details

Definition of automated decision system

For the purposes of the ADS Workgroup, the following definition of an automated decision system shall be used:

- **"Automated decision system"** or "system" means any algorithm, including one incorporating machine learning or other artificial intelligence techniques, that uses data-based analysis or calculations to make or support government decisions, judgments, or conclusions that cause a Washington resident to be treated differently than another Washington resident in the nature or amount of governmental interaction with that individual including, without limitation, benefits, protections, required payments, penalties, regulations, timing, application, or process requirements.

Authorizing statute

The automated decision systems workgroup is authorized in law by Chapter 334, Laws of 2021, Sec. 151(14).

Contact

Questions regarding the ADS Workgroup or this charter can be directed to:

- Katy Ruckle, J.D., State Chief Privacy Officer (privacy@ocio.wa.gov)
- Derek Puckett, Legislative Affairs Director (derek.puckett@watech.wa.gov)

Appendix A: ADS Research Findings

Governmental entities are increasingly using automated decision systems to automate or support both simple and complex decision-making processes.[6] These systems use algorithms, or a series of steps, to transform inputs into outputs that support or make decisions. These algorithmically driven systems may be used to allocate healthcare services and benefits, facilitate urban planning, evaluate teachers, detect unemployment fraud, forecast crime, surveil individuals, or determine the length of someone's sentence.[7]

Automated decision systems are often seen to reduce costs, improve delivery of public services, and make decisions more efficient, reliable, and accurate. In some cases, use of automated decision systems may be mandated in legislation. However, a growing body of evidence indicates that automated decision systems are often discriminatory, unreliable, and inaccurate; lack transparency and accountability; and may undermine the legitimacy of the governmental entities that use them.[8]

States such as Arkansas and Idaho that built automated decision systems to automate benefits distribution caused immense harm to thousands of disabled Medicaid recipients, whose essential care needs were drastically cut by non-transparent and unaccountable algorithms.[9] Similarly, jurisdictions such as Broward County, Florida are using or have used algorithmic systems to predict recidivism rates, though such systems have exhibited significant bias against Black defendants.[10] The discrimination and harm that can be caused by automated decision systems have been extensively documented in areas including but not limited to policing, education, child welfare, housing, health care, and credit.[11]

Washington state agencies currently deploy a large number and range of automated decision systems—some of which use simple rules-based algorithms (e.g., some tools that automate paper and pencil decision-making tools) and others that incorporate machine learning. It is important to note that the benefits and risks posed by use of these systems depend on the specific social, political, and institutional contexts in which they are deployed. Even simple rules-based algorithmic systems can pose significant risks if they affect a large number of people or if they support high-stakes decisions. More complex machine-learning based systems can introduce additional risks such as the effects of being trained on biased or discriminatory data.

In recent years, there has been a rapid evolution in the understanding of how algorithmically driven automated decision systems operate as well as the risks posed by their use. Some systems in use

---

[6] https://ainowinstitute.org/nycadschart.pdf

[7] https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/

[8] https://scholarlycommons.law.emory.edu/cgi/viewcontent.cgi?article=1418&context=elj

[9] https://cdt.org/wp-content/uploads/2020/10/2020-10-21-Challenging-the-Use-of-Algorithm-driven-Decision-making-in-Benefits-Determinations-Affecting-People-with-Disabilities.pdf

[10] https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

[11] https://par.nsf.gov/servlets/purl/10074337

by governmental agencies today have not been audited for biases, and in many cases, were developed several years ago when techniques for identifying and addressing biases were not as advanced as they are today. An increasing number of jurisdictions are regulating automated decision systems (see ADS landscape section), understanding that while governmental entities may seek the benefits that these systems purport to deliver, there are substantial risks of using such systems that must be addressed to prevent harm. A consensus is beginning to emerge on some basic principles (see Guiding Principles section in Appendix B).

Washington state has introduced one of the country's first pieces of legislation to consider application and use of automated decision systems. The state can be a leader in ensuring that government use of automated decision systems does not cause discrimination and other types of harm.

Goals of using automated decision systems:

- Efficiencies: Automated decision systems may assist governmental entities in streamlining processes and making administrative decisions more quickly and efficiently.
- Cost savings: Automated decision systems may reduce costs, by replacing or reducing the number of human decision-makers or by reducing the time necessary to record data collected, conduct calculations, and make decisions.
- Reducing the biases and inaccuracies of current systems: While automated decision systems may introduce and reinforce biases, they may also potentially provide an opportunity to address them.
- Improving delivery of public services: Automated decision systems may help the public receive improved and more accessible services.

ADS may provide an opportunity for significantly improved transparency in how decisions are being made.

Because ADS operate by a defined algorithm and make decisions consistently, there are broad opportunities to test their impact and make improvements over time.

Risks of not using automated decision systems:

- Many automated decision systems used in state government today were put in place many years ago to replace paper and pencil systems that relied more heavily on individual human judgement. In the absence of the ability to automate at least some of these basic service delivery decisions, agencies may need to revert to human-based systems that are themselves subject to other types of human variation and bias.

Risks of using automated decision systems:

- Discrimination: Automated decision systems may reproduce existing patterns of discrimination that are already present in our society.
- Inaccuracies: Automated decision systems reproduce and could exacerbate existing inaccuracies and biases and may be just as or more inaccurate than human decision-makers.

- Automation Bias: Humans tend to place too much trust in automated decisions and their suggestions and ignore contradictory information made without automation. Humans may overestimate the accuracy of decision support and decision-making systems that may be as or more error-prone than human decision-makers. Automation bias may obfuscate and exacerbate biased and inaccurate decision-making.[12]
- Non-transparency: It is difficult or impossible for individuals to know if an automated decision system is being used, how that system operates, and the impacts of the system on individuals and society.
- Lack of explainability: Some automated decisions systems are very difficult to explain in clear and concise language that would be understandable to those auditing the system or those potentially impacted by their use. This risk may be especially prevalent when an automated decision tool is procured through a third-party vendor.
- Lack of accountability: Individuals who are affected by automated decision systems may not have the ability to meaningfully challenge a system's decisions. Governmental entities that adopt automated decision systems may not consult the individuals and communities that may be affected by use of those systems and may not have a human-centered dispute resolution process.
- Threats to privacy: Large amounts of data about individuals are often used to train automated decision systems and used by both simple and complex automated decision systems to transform inputs into decisions or suggestions. Individuals may not understand or have given consent for their data to be used for such a purpose. Additionally, many automated decision systems are surveillance tools, such as facial recognition technology.
- Threats to legitimacy and public trust: Use of automated decision systems undermine the legitimacy and public trust of governmental entities when such entities re-delegate their decision-making responsibilities to unaccountable and nontransparent systems.

To mitigate the risks of using automated decision systems, many jurisdictions are, at a minimum, adopting the following practices:

- An evaluation of whether it is appropriate for governmental entities to use the automated decision system at all. This evaluation should incorporate public participation and comment.
- Independent testing for bias and inaccuracy.
- Transparency so that the public may know that algorithmic systems are being used in the first place, they may assess information about the systems, and may demand responses about their use.
- Accountability so that the public may meaningfully challenge a system's decisions.

Additionally, in the report, "Algorithmic Accountability for the Public Sector," (August 2021), The Ada Lovelace Institute, AI Now Institute, and Open Government Partnership examined the

[12] https://arc.aiaa.org/doi/10.2514/6.2004-6313

implementation of 40 different algorithmic accountability policies by more than 20 national and local governments and identified the following key findings.[13]

- Clear institutional incentives and binding legal frameworks can support consistent and effective implementation of accountability mechanisms, supported by reputational pressure from media coverage and civil society activism.
- Algorithmic accountability policies need to clearly define the objects of governance as well as establish shared terminologies across government departments.
- Setting the appropriate scope of policy application supports their adoption. Existing approaches for determining scope such as risk-based tiering will need to evolve to prevent under- and over-inclusive application.
- Policy mechanisms that focus on transparency must be detailed and audience appropriate to underpin accountability.
- Public participation supports policies that meet the needs of affected communities. Policies should prioritize public participation as a core policy goal, supported by appropriate resources and formal public engagement strategies.
- Policies benefit from institutional coordination across sectors and levels of governance to create consistency in application and leverage diverse expertise.

[13] See ADS landscape section and pages 10-16 of the "Algorithmic Accountability for the Public Sector" report for more details. https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/

Appendix B: Guiding Principles for Governmental Entities Using Automated Decision-Making Systems

During its work, the ADS Workgroup heard from several experts on the procurement, implementation, and management of ADS and reviewed articles on those issues. In those presentations and articles, several themes emerged on what governmental entities should aspire to in the procurement, implementation, and management of ADS. Those themes are referred to herein as "Guiding Principles." Although there may be some differences among experts, and not a consensus among ADS Workgroup members, the themes of experts are fairly consistent across the spectrum of advice.

It is the work of the Legislature and the Executive branch of governments to determine whether and how to use the Guiding Principles in the context of legal constraints and available resources. Several ADS Workgroup members have expressed concern that the elements of Guiding Principles will be taken as uniform mandates to be implemented immediately and completely. That is not the intent.

The ADS Workgroup's Recommendations follow from the Guiding Principles and legal and resource considerations. Overall, the ADS Workgroup accepts that the practical and policy considerations behind the procurement, implementation, and management of ADS systems are best made by keeping the Guiding Principles in mind. The Guiding Principles should be viewed as the goals and key considerations to be used by governmental entities procuring, implementing, and managing their ADS. Weighing how and when to translate goals into reality is the important work of government.

1. Prioritization: The governmental entity should develop a prioritization framework for allocating resources to address existing and future ADS. The prioritization framework could include criteria such as whether the system: (1) creates significant effects on identified or identifiable natural people, (2) affects a large number of natural people, or (3) involves a high risk of error, such as systems that have been developed without transparency of the information used to develop the algorithm or specific content of the algorithm created, or without any testing for bias or inaccuracy; (4) the level of discretion the agency has in creating the algorithm and (5) the level of automation and opportunity for human review. A sample prioritization matrix is included below. Agencies should avoid using the magnitude of effort involved as a reason for not examining a priority system. While a final prioritization framework is being developed, agencies should adopt and implement an interim prioritization framework so as not to delay work relating to ADS. The governmental entity should make its prioritization framework available to the public where the ADS has a significant effect on natural people.
2. Evaluation Whether to Adopt System: Prior to procuring, developing, or using an ADS, the governmental entity (the utilizing agency or an independent entity) should assess the reasons for adoption of the systems and the risks involved in its design and implementation. These criteria should utilize or maximize the existing procurement process as much as possible to

attain these goals to benefit limited tax dollars. That assessment would normally include information on:

a. The benefit of the system in fulfilling the agency's mission and societal goals.

b. Testing to determine whether the system or the sources of data used in the design appear to create inaccuracies, bias or disproportional effects and if so, steps taken to address those effects.

c. Whether the system is secure against wrongful data disclosure or malicious manipulation.

d. Whether use of an ADS for the decisions being made will adversely affect the public's trust in Washington state governmental action.

e. The opportunity for participation by the public, including notifying the public about the risks and benefits of the system and giving individuals and communities affected by the use of ADS, meaningful input in the decision on whether it should be used and how it is designed.

f. Documentation for future review of the process and decisions that is specific and transparent to the public.

g. Extreme caution in adoption of any new system where the development did not include testing for bias or inaccuracy with disclosures such that the validity of the testing is subject to independent review by members of the public or independent entities.

3. Updated Assessments: The assessment of the ADS utilized in procurement or development should be updated and re-assessed during any subsequent modifications of the ADS and/or to the data collection process used to inform the algorithm. These updates should be transparent to the public.

4. Periodic Implementation Testing: There should be a process (preferably a third-party audit) for periodically determining, during the use of the ADS, whether the system appears to create inaccuracies, bias, or disparate outcomes. Any identified tendencies toward inaccuracies or bias should be addressed or the reasons for not addressing them set forth with specificity. This process and the resulting report should be transparent to the public. This testing for priority systems is extremely important to ensure fairness of governmental use of the system.

5. Transparency of the Algorithm: Absent compelling reasons to the contrary, the ADS's algorithms and source code should be released to make the actual algorithm available for public scrutiny. Examples of a compelling reason may include the creation of a factually based demonstrable threat to governmental integrity or the creation of a substantial risk that an individual will circumvent the system in a way that would create a factually-based demonstrable threat to governmental integrity or a material cybersecurity risk. Systems that restrict release of the algorithm or source code for commercial reasons should be subject to particularly rigorous third-party evaluation for risk of inaccuracy or bias. Further, examination of the algorithm/source code is important but is no substitute for actual testing for inaccuracy or bias in design and implementation of ADS.

6. **Audit Trails:** The ADS should generate audit trails that record the facts and rules supporting its decisions. This will allow the governmental entity to provide natural persons with the reasons supporting an ADS's adjudication and will facilitate eventual judicial review of the decisions. This information should be made available to third-party researchers, allowing for independent investigation of the system's accuracy and bias.
7. **Training on Automation Bias:** Persons involved in procurement, development or operation of ADS should receive explicit training about the phenomenon of automation bias.
8. **Evaluation of Risks/Determination Whether to Proceed:** Prior to use of the ADS, and at any time a risk of inaccuracy or bias is discovered, the governmental entity should consider whether the risks and effects of inaccuracies or bias on natural persons and potential loss of the public's trust are at such a magnitude that the system should not be used at all. Any decision in this regard should consider input from the persons who will be affected by the system, and the ultimate decision should be specific and transparent to the public.
9. **Review of Decisions by those Affected:** Those affected by a decision made or supported by an ADS should be able to review and challenge the basis for that decision, particularly where significant rights or responsibilities of natural persons are involved.
10. **Weighing Advantages Against Known Bias or Inaccuracies:** Although benefits created by ADS systems are important, in weighing the advantages of benefits against harm to natural persons by known or likely bias or inaccuracies, the prevention of harm to natural persons should be given substantially greater weight.
11. **Review of Current Systems and Processes/Necessary Action:** Using the prioritization framework (first an interim framework and then a finalized one) to identify ADS for examination, the governmental entity should determine the extent to which the procurement and implementation of its ADS has complied and is complying with these Guiding Principles. Where the procurement or implementation of a system is not in compliance with the Guiding Principles, the governmental entity should identify the reasons for that non-compliance and the appropriate action in accordance with prioritization principles.

Sample Prioritization Matrix

Qualifying ADS could be assessed (within a time period) and results published (taking account of limitations due to security, etc.). The reason for the rating decision by the agency could be explained in clear language along with context information such as a description of decision being made, the approximate size of impacted population, and how long it has been in use.

The risk assessment could help agencies prioritize bias assessment on existing and potential procurements.

Possible risk of bias rating [for example]:

Effect on natural people

- **Low**: Decision does not impact legal rights or the provision of services or scrutiny that could lead to an impact on legal rights or services.
- **Medium:** Decision impacts processing, relatively minor services or legal rights or financial impacts on individuals. Small number of impacted clients.
- **High:** Decision can have a significant impact on the provision of services, financial impact, or legal rights. Large number of impacted clients.

Likelihood of bias

- **Low:** Decision directly follows federal or state regulations or follows adopted policy or rule.
- **Medium:** Developed with disclosure of information used and the algorithm has been tested for bias and inaccuracy.
- **High:** Developed without disclosure of information used or the algorithm created and has not been tested for bias or accuracy.

Complexity

- **Low:** simple decision rule.
- **Medium:** simple calculation of existing data elements (i.e., a weighted average).
- **High:** complex algorithm, estimation, machine learning, etc.

Some matrixes could be used to figure out a final rating such as:

	Low impact	Medium impact	High impact
Low likelihood	5	4	4
Medium likelihood	4	3	3
High likelihood	4	2	1

Complexity could be used to determine the type of review, for example:

- **Low:** ACLU questions (See Appendix C).
- **Medium:** Outcome analysis.
- **High:** Assessment by professional or third party.

Appendix C: Complete Questions and Answers regarding WA ONE System

Automated Decision System Budget Proviso Implementation

Questions for Agencies from ACLU as approved by ADS Workgroup

1. What is the automated decision system's name, vendor and version?
   - Washington Offender Needs Evaluation (WA ONE).
   - Vendor – procured in 2015; Dr. Zachary Hamilton and WSU.
   - Ongoing maintenance and updates performed in-house at DOC.
2. What decisions is the system utilized to make?
   - Identifies risk level classifications for all individuals under DOC jurisdiction. Not just those housed at DOC facilities.
   - Program priorities are identified – what an individual's domain needs are. (i.e. high substance abuse need prioritized for substance abuse treatment).
   - Used to make determinations concerning risk-based supervision.
   - The output need domains used for decision making: Aggression, Attitudes/Behaviors, Education/Vocational, Employment, Mental Health, Residential, Social Influences, Substance Abuse – High/Mod/Low score.
3. What data is input into the system?
   - Approximately dynamic manually input responses as well as data from system (106 items, 20 static, 9 semi-dynamic, 77 dynamic).
   - Questions on the following domains: Demographics, Juvenile Record, Adult criminal history, correctional events, residential status, education, vocational(work) status, employment status, relationship status, family status, children, friends, leisure time, alcohol/drug use, mental health (including hospitalization & medications), aggression, attitudes/behaviors.
   - Criminal History + Needs Interview + Correctional Events = Washington ONE Recidivism Model.
4. How is the input data gathered, how often is it updated, and are subjective inputs ever audited for consistency across data collectors?
   - Mix for input data gathered. OMNI – Offender Management Network of Information.
   - Gathered by counselors via interview with the individual. "Motivational interview" to elicit responses from the person being interviewed. Conversational – to gather narrative and context.

- Reassessment schedule – when the conviction record is updated. But can occur at any time based on new information. At intervals based upon time to Eligible Release Date (ERD) or Supervision End Date (SED).
    i. Criminal Conviction Record update (within 30 days).
    ii. Every six months to a year based on ERD/SED.
    iii. Within 30 days before transfer to partial confinement.
    iv. Between 60 and 90 days before release to the community from a work/training release.
    v. When required as part of a quality of review process.
- Inter-rater reliability not automated at this time re subjective inputs.
- Audited initially during development.
- There is a QA performed on assessments – not accuracy across. Supervisor review for accuracy.
- Reviews done on all pre-closure cases.
- Released – low or moderate audit process performed by correction specialist.

5. Is the decision algorithm available for examination by the agency and/or the public?
    - Yes. for the agency to a certain extent, but not the public.
6. Has there been any public or community engagement used in selection or design of the system? If so, please describe this engagement.
    - There was a bid process that was posted publicly.
    - There has been engagement with the Family Council about the Washington ONE system to answer questions about the assessment.
    - Pre-implementation presentations and consultations with legal counsel.
7. Does law or regulation mandate any of the decision system criteria? If so, which criteria?
    - Yes, DOC is required to use the risk assessment system endorsed by WSIPP. See RCW 72.09.270.
8. Do the system's decisions intentionally differentially affect members of protected classes, such as selecting persons with disabilities for certain benefits?
    - Yes, differentiates by gender to allow tool to assign supervision levels. System is designed to be gender responsive to consider differential needs based on gender in the algorithm. Intended to provide for needs of the group by also identifying case management relevant issues such as ADA, behavioral health, medical and other needs that may indicate a need for supportive activities.

9. Has the system been tested for unintended bias by the agency or an independent third party? If so, what were the results? Describe briefly the nature of the testing.
    - Yes, during system's development it went through initial analyses of biases that were considered in design.
    - WSIPP contract to evaluate current state of potential of bias based on demographic.
    - Another analyses of bias post-*Blake* decision current state.
    - Norming period after tool launch considering potential bias and making adjustments.
10. Has the system produced known erroneous results and if so, what were those errors (including the results of any audits conducted to check for erroneous results)?
    - No known erroneous results or major audits beyond quality assurance (QA) work.
    - Known data entry issues which are identified and corrected through QA process.
    - Norming period report in Oct. 2020. Published by WSIPP.
    - System went live Dec. 15, 2017.
11. In addition to any intentional differential effect on members of a protected class, are there are other differential effects on protected classes as shown by comparison of the system's data to general census data or, where relevant, subpopulation data, such as the effect on justice system defendants of color as contrasted with all defendants? If audits have been performed to determine such differential effects, what were the results of those audits?
    - Yes, some differential effects are expected (especially for the needs domains which determine treatment planning) depending on the protected class. E.g., someone with a history of mental and behavioral health issues/diagnoses is going to score as high needs in Mental Health. Mental/behavioral health issues is a high correlate with many co-occurring disorders/disabilities).
    - Census data as a benchmark for fairness or bias is unreliable.
    - Other differential effects are due to criminal justice system as a whole - e.g., Age curves.
    - Same response as number 9.
    - DOC contracting with third party to analyze again and in the future.
12. Can those affected by a system decision review and challenge the basis for that decision? If so, how, and what were the results of any such challenges?
    - Yes, individuals can challenge the results of an assessment.
    - It is in DOC policy.
    - Process goes to superintendent or field administrator.
13. Is the decision system operated by a third party? If so, what rules govern such operation and what audits are conducted to ensure compliance?
    - No, system is not operated by a third party.

14. What is the fiscal impact of the system, including initial cost, operating costs, and any cost savings established as flowing from use of the system?
    - All costs not readily available. RFP – 2015 Work started in 2016.
    - Mandated by law to have the system.
15. What were the personnel hours required to gather the relevant information (questions 1-14) for the system examined?
    - 15.5 hours.

Appendix D: ADS legislation, regulation, and policy in other jurisdictions

As governments turn to algorithms to support public decision-making, legally binding commitments, high-level principles, and voluntary guidelines are being issued in the United States and all around the world to ensure algorithmic accountability. This appendix provides a summary of key initiatives. While this is only a snapshot of the landscape at a point in time, we hope that the perspectives are useful as Washington state considers its own policies.

In the US, most legislation to date has been in the form of prohibiting specific uses of ADS:

- Governmental use of facial recognition has been prohibited in King County, Washington state's most populous county; Portland, Oregon; the state of Maine; and over a dozen other cities around the country.
- Vermont, Virginia and several cities have banned law enforcement use of facial recognition (while continuing to allow other governmental uses).
- Bellingham, Washington, Santa Cruz and Oakland, California have banned predictive policing.

Several states in addition to Washington are also considering state-level legislation, for example banning discrimination by ADS, requiring an inventory of ADS systems, or introducing guidelines for governmental ADS systems.

In 2018, New York City passed a law establishing an automated decision systems task force to provide recommendations on sharing information about ADS with the public and how agencies may address instances where people are harmed by agency ADS. The 2019 report, and the Confronting Black Boxes Shadow Report endorsed by two dozen AI experts and a range of civil rights and algorithmic justice organizations, both feature many recommendations relevant to other jurisdictions as well.

At the federal level, the White House Office of Science and Technology (WHOSTP) has just announced the development of a Bill of Rights for an AI-Powered World. While the details are still to be determined, the leaders of WHOSTP note that possible rights could include "freedom from being subjected to AI that has not been carefully audited to ensure that it's accurate, unbiased, and has been trained on sufficiently representative data sets": and "a right to meaningful recourse if the use of an algorithm harms you." Outcomes could potentially include the federal government refusing to buy software or technology products that fail to respect these rights or requiring federal contractors to use technologies that adhere to this "Bill of Rights."

Elsewhere, Canada's 2019 Directive on Automated Decision-Making requires an Algorithmic Impact Assessment (AIA) to determine the impact level of a system. Additional requirements include **testing for biases and third-party review**; **meaningful explanations** of decisions and a plain language description of how the system works; **the right to access and test the software** and to authorize external parties to review and audit these components as necessary; and **open-source release of any software developed by the government**.

Europe has moved even farther towards broad regulation of ADS. Most important is the draft European Union (EU) Regulation on Artificial Intelligence (also known as the "Draft AI Act"). This act:

- Prohibits AI systems with "unacceptable risk", including "social scoring" AI systems used by public authorities to generate 'trustworthiness' scores and lead to unjustified or disproportionate treatment of individuals or groups; manipulative systems likely to cause harm that use subliminal techniques or exploiting vulnerabilities due to age or disability; and some uses of real-time biometric systems (including facial recognition) in publicly accessible spaces by law enforcement, with exceptions.
- Introduces requirements and obligations on "high risk" systems related to access to and enjoyment of essential services and benefits; law enforcement; biometric identification and categorization; and several other categories. Requirements include a **risk management system**; **data quality criteria**; **accuracy, robustness, and cybersecurity**; and **building for human oversight**.

A revised version of this act is expected to be approved in 2022. In October 2021, the European Parliament adopted the (non-binding) Report on Artificial Intelligence in criminal law and its use by the police and judicial authorities in criminal matters, whose recommendations largely align with the Draft AI Act.

The August 2021 report "Algorithmic Accountability for the Public Sector" (by the Ada Lovelace Institute, AI Now Institute, and Open Government Partnership) surveys algorithmic accountability policies around the world. As well as a useful classification of types of policies -- including **principles and guidelines**, **procurement conditions** and **oversight bodies** as well as prohibitions and AIAs -- this report also includes learning from what the authors characterize as "the first wave" of regulation.

The remainder of this section has additional details on the initiatives briefly summarized here.

In the United States

a) Federal level

In October 2021 the White House Office of Science and Technology Policy announced the upcoming development of a Bill of Rights for an AI-Powered World, intended to require powerful technologies to respect America's democratic values and abide by the central tenet that everyone should be treated fairly. According to Alondra Nelson and Eric Lander, leaders of the Office, the rights and freedoms they expect data-driven technologies to respect could include:

- A right to know when and how AI is influencing a decision that affects your civil rights and civil liberties.
- Freedom from being subjected to AI that has not been carefully audited to ensure that it's accurate, unbiased, and has been trained on sufficiently representative data sets.
- Freedom from pervasive or discriminatory surveillance and monitoring in your home, community, and workplace.

- A right to meaningful recourse if the use of an algorithm harms you.

Likewise, to effectively protect these rights, measures could include:

- The federal government refusing to buy software or technology products that fail to respect these rights.
- Requiring federal contractors to use technologies that adhere to this "Bill of rights."
- Adopting new laws and regulations to fill gaps.

b) State-level

In October 2020 the Vermont Legislature enacted S.124, which in Sec. 14 sets forth a moratorium on the use of facial recognition technology by law enforcement officers. This, until the Vermont Criminal Justice Council recommends a statewide policy on law enforcement officers' acquisition and use of facial recognition technology, which considers any law enforcement needs to use facial recognition technology, as well as any potential inaccuracies or other limitations in the capacities of that technology, including implicit biases, and an opportunity for community involvement and feedback.

In July 2021 Maine's legislature unanimously passed LD 1585, which prohibits the use of facial recognition technology in most areas of government, including in public schools, and for surveillance purposes. This prohibition applies across all levels of state, county, and municipal government, with extremely limited exceptions.

Besides these two statewide bills, the following states have pending legislation in relation to ADS:

- California:
  - CA A.B. 13: This bill would require:
    - The Department of Technology to:
      - Establish and make public guidelines for identifying automated decision systems that are subject to the bill's requirements.
      - Conduct a comprehensive inventory of all high-risk automated decision systems that have been proposed for, or are being used, developed, or procured by state agencies.
      - Repeat this inventory and report process in 2025 and 2027.
      - Develop a sample automated decision system impact assessment report for prospective contractors.
    - State agencies:
      - To encourage a bid response submitted by a prospective contractor of goods or services that include the use, licensing, or development of an automated decision system for a high-risk application to include an automated decision system impact assessment report that makes specified disclosures.

        - To submit to the department a high-risk automated decision system impact assessment report, if included in the bid response.

- Massachusetts:
    - MA H.119: This bill would establish a commission within the executive office of technology services and security for the purpose of studying and making recommendations relative to the use by the commonwealth of automated decision systems that may affect human welfare, including but not limited to the legal rights and privileges of individuals.
    - MA S. 60: Similar to the previous one, this bill would establish a commission on transparency and the use of artificial intelligence in government decision-making.
- New Jersey:
    - NJ S.B. 1943: This bill would prohibit a person, bank, banking organization, credit reporting agency, mortgage company, or other financial institution, lender, or credit institution involved in the making or purchasing of any loan or extension of credit to discriminate through the use of an automated decision system against any person or group of persons who is a member of a protected class.
- Vermont:
    - VT H.B. 263: This bill would prohibit any automated decision system to discriminate against any individual based on race, color, religion, national origin, sex, sexual orientation, gender identity, ancestry, place of birth, age, crime victim status, or physical or mental condition. Also, it would require the Secretary of Digital Services to:
        - Conduct a review and make an inventory of all automated decision systems that are being developed, used, or procured by the State.
        - Adopt standards and practices regarding State development, procurement, use, and issues of bias, in automated decision systems.
    - VT H.B. 429: This bill would establish an advisory committee to address bias in software programs used by the State, in charge of creating anti-bias standards.
- Washington:
    - WA S.B. 5116: This bill would establish guidelines for government procurement and use of automated decision systems to protect consumers, improve transparency, and create more market predictability.

c) Local level

In 2018, New York City passed Local Law 49 of 2018, which established an automated decision systems task force that would provide recommendations on how information on agency automated decision systems may be shared with the public and how agencies may address instances where people are harmed by agency automated decision systems. After 18 months of work, in November 2019 the task force submitted to the mayor and the speaker of the council a report with the

following recommendations (key general principles where the task force members reached consensus):

- *Develop and centralize resources within the City government that can guide policy and assist agencies in the development, implementation, and use of ADS:* This includes establishing a centralized ADS Organizational Structure within City government; incorporating key principles of fairness, transparency, innovation and efficiency, and accountability to help guide responsible City agency use and management of ADS, and advising agencies on compliance with laws or regulations that may affect their use of ADS.
- *Adopt a phased approach to developing and institutionalizing agency and citywide ADS management practices:* This includes creating a framework for identifying ADS that should be prioritized, considering the general descriptive characteristics of the system, the explainability of the tool, the urgency of the need for the tool or system and the downsides of a delay in, or not implementing the tool or system, and its impact, and incorporating flexibility into management processes.
- *Strengthen the capacity of City agencies to develop and use ADS:* This includes providing agencies with sufficient funding and staffing for ADS management; considering agency expertise in developing centralized policy regarding ADS management; creating best practices on ADS, including ADS procurement, data retention, and data sharing, to serve as a resource for agencies; educating agency staff on ADS and how to communicate about ADS with the public, and enabling the input of experts external to the City to help support the City's work of ADS management.
- *Facilitate public education about ADS:* This includes creating a visible, accessible presence for ADS management; develop educational materials in plain language; helping individual City residents request additional information from agencies about ADS and reporting to the public about the overall ADS management.
- *Engage the public in ongoing work around ADS:* This includes providing an opportunity for public input to the Organizational Structure's guidelines for agencies on ADS management and involving impacted communities in discussions about specific ADS.
- *Establish a framework for agency reporting and publishing of information related to ADS:* This includes identifying the highest-priority tools; establishing reporting standards for information related to ADS; publishing agency reported information about ADS where legally permissible and aiding compliance with laws and regulations.
- *Incorporate information about ADS specifically, where relevant, into processes for public inquiry about or challenge to City agency decisions:* This includes integrating general ADS information into preexisting inquiry response channels; providing guidelines to agencies on how to respond to and document specific public inquiries and challenges and creating a single point of contact in the City for individuals to submit questions or comments about specific ADS decisions.
- *Create an internal City process for assessing specific ADS for any risk of disproportionate impact to any individual or group on the basis of protected characteristics*: This includes providing guidelines on which ADS should be subject to review; developing options for

protocols for assessment; providing opportunities for impacted communities and others from outside the City to provide input and developing a process for responding to instances of negative disproportionate impact on the basis of protected characteristics.

Confronting Black Boxes, the Shadow Report (written by Rashida Richardson, who is now at WHOSTP, and endorsed by a wide range of civil rights, immigrant rights, civil liberties, and algorithmic justice organizations) also has some excellent recommendations covering topics like privacy, archiving, public notice, and addressing instances in which a person are harmed by agency ADS. The Shadow Report also has recommendations for tasks forces and legislators in other jurisdictions, for example:

- Any government body or process that is empowered to review, assess, and make recommendations regarding government agencies using automated decision systems.
- Must have budgetary support to compensate or support participation of non-civil-servant members, and hire independent experts or consultants needed to support the mandated process.
- Legislation permitting use or funding pilots of automated decision systems should require periodic reviews and/or retrospective studies to assess the accuracy of the system, potential bias or disparate outcomes, and the value of the system in fulfilling the agency's mission and societal needs or interests

Also, and among many other cities in the United States, in June 2020 Santa Cruz, California issued Ordinance No. 2020-17, by which it prohibited the city's acquisition and/or use of face recognition technology and predictive policing technology, prior to obtaining City Council approval, by resolution, based on the City Council's finding that the technology meets scientifically validated and peer-reviewed research, protects and safeguards the civil rights and liberties of all people, and will not perpetuate bias.

In January 2021 Oakland, California followed suit, issuing an Ordinance that prohibited the city's acquisition and/or use of Biometric Surveillance Technology and Predictive Policing Technology.

Similarly, in June 2021 Seattle-area King County passed Ordinance 19296, prohibiting the acquisition and use of facial recognition technology by County administrative offices and executive departments, including the department of public safety.

Most recently, in November 2021 the City of Bellingham passed Initiative No. 2021-02, prohibiting city from acquiring or using facial recognition technologies.

Other cities banning facial recognition include San Francisco, Berkeley, and Alameda in California; Somerville, Cambridge, Northampton, Springfield, and Boston in Massachusetts; Minneapolis; Jackson, Mississippi; New Orleans; Portland, Maine; and Portland, Oregon.

Around the world

d) 20 National and local jurisdictions in Europe and the United States.

In the report "Algorithmic Accountability for the Public Sector" published in August 2021, the Ada Lovelace Institute (Ada), AI Now Institute (AI Now), and Open Government Partnership (OGP) synthesize the different mechanisms included in more than 40 algorithmic accountability policies

being drafted or implemented in more than 20 national and local jurisdictions in Europe and the United States. According to the report, the kinds of policy responses from governments towards algorithmic accountability can be classified in the following eight policy mechanisms:

1. Principles and guidelines: documents that identify high-level policy goals, and how they might be implicated in the use of algorithmic systems by public agencies. Are generally not intended to be binding and are usually issued as normative standards against which agencies can assess their use of algorithmic systems. These standards include guiding principles for the use of AI in government such as "general interest," "respect for human rights," "transparency" and "privacy by design."

   Examples of policies that include this mechanism are (i) Uruguay's AI Strategy for Digital Government; (ii) Australian Ombudsman's Better Practice Guide on Automated Decision-Making; and (iii) UK's Data Ethics Framework.

2. Prohibitions and moratoria: orders that ban or prohibit the use of particular kinds of 'high risk' algorithmic systems, where the perceived risk or harm, in a specific context, is considered to be too high to justify its use. Have been most prominently applied to facial recognition technologies used by law enforcement, and in some cases, by local governments. In some cases, these prohibitions are expressly time-limited and are framed as temporary moratoria, until accountability frameworks are implemented.

   Examples of policies that include this mechanism are (i) bans on the use of facial recognition technology ('FRT') in San Francisco, Oakland, and Seattle; and (ii) Morocco's facial recognition policy.

3. Public transparency: require the publication and/or provision of information about algorithmic systems to the general public, usually through public registries of algorithmic systems, requirements for source code transparency, or explanations of algorithmic logics. Even, some frameworks require not only information about the technical components, but also about how data is collected, stored, and secured, or about the administrative processes behind the system. In all mechanisms reviewed, the transparency requirements are subject to exceptions owing to countervailing policy objectives such as trade secrets, system security concerns, or privacy.

   Examples of policies that include this mechanism are (i) Article L-312-1-3 of the French Digital Republic Bill; (ii) the Canadian ADM Directive; (iii) the Aotearoa New Zealand Algorithm Charter; (iv) UK's Data Ethics Framework; and (v) some form of an algorithm registry implemented in Ontario, Amsterdam, Helsinki, and in cities in France, including Antibes, Lyon, and Nantes.

4. Algorithmic impact assessments (AIAs): policy mechanism usually conducted prior to the actual 'live' usage of algorithmic systems or concurrent to the use of such systems. Are intended to define and construct a matrix of harms, benefits, and risks, to evaluate ex-ante whether the deployment of an algorithmic system is suitable in a particular context, and if not, what measures must be taken by a responsible actor to respond to the possibilities of harm. In some cases, the outcomes of AIAs go on to determine the eventual level of regulatory scrutiny applied to algorithmic systems.

Examples of policies that include this mechanism are (i) the Canadian Directive on Automated Decision-Making; (ii) the Aotearoa New Zealand Algorithm Charter; and (iii) Tamil Nadu Safe and Ethical AI Policy.

5. Audits and regulatory inspection: a range of practices for inspecting the working of a particular algorithmic system, to understand its functioning, and assess it with respect to some predefined normative standard. Are generally carried out after, or concurrent with, the use of a system, and tend to be conducted by a third, second, or first party to the audited organization. They can be of two types: (a) a technical audit: examines the technical elements (inputs, outputs, algorithms) to assess reliability, check for discriminatory biases in results, or assess other aspects of the functioning of the algorithmic system; and; (b) a regulatory inspection: examines the functioning of an algorithm system, with reference not only to its technical element, but also with a focus to assess it against some normative standard (for quality assurance, legality, etc.).

   Examples of policies that include this mechanism are (i) the Draft Guidance on AI Auditing released by the UK Information Commissioner's Office (ICO); and (ii) the ad hoc audits conducted by the National Audit Office of Sweden, the Netherlands Court of Audit, the UK Centre for Data Ethics and Innovation (CDEI), and the County of Allegheny, Pittsburgh.

6. External/independent oversight bodies: bodies in charge of overseeing and directing the use of algorithmic systems by public agencies. Are typically responsible for monitoring the actions of a public agency and making recommendations, sanctions, or decisions about their use of algorithmic systems. In some cases, they are expected to provide non-binding guidance on issues including human-rights compliance, scientific validity, privacy, and ethics. Finally, oversight bodies can also act as forums where diversity of expertise and participation can be brought together.

   Examples of this type of mechanism are (i) the New York City Automated Decisions Task Force; (ii) the task forces included in the Community Control of Policy Surveillance (CCOPS) legislation in Seattle and Oakland; (iii) the New Zealand Data Ethics Advisory Group; and (iv) the West Midlands Police Data Ethics Committee.

7. Rights to explanation, hearings, and appeal: procedural protections intended to implement fair processes and provide forums for affected individuals or groups to debate or contest particular decisions that affect them. They look to either allow affected individuals and groups to contest decisions that do not meet specific legal standards or to introduce human agency into an algorithmic process and identify particular persons as having ultimate responsibility for decisions taken with the use of algorithmic systems. As so, they can include: (a) providing notice of the decision and a hearing to the affected parties; (b) the duty to provide reasoned decisions and explanations of a decision; (c) the right of affected parties to present evidence; and/or (d) the right to have "human intervention" (human-in-the-loop) in the decision-making process.

   Examples of policies that include this mechanism are (i) the EU General Data Protection Regulation (GDPR); (ii) the Canadian Directive on Automated Decision-Making; and (iii) the Aotearoa New Zealand Algorithm Charter.

8. Procurement conditions: rules governing the procurement and acquisition of algorithmic systems by governments and public agencies, that limit the design and development of such systems. By establishing contractual pre-conditions for acquiring algorithmic systems, they ensure that certain transparency and accountability standards are met.

   Examples of policies that include this mechanism are (i) the Tamil Nadu Safe and Ethical AI Policy; and (ii) the standard clauses for the procurement of algorithmic systems of the City of Amsterdam.

In addition to this typology, the "Algorithmic Accountability for the Public Sector" report identifies the following six lessons that can be concluded from the reviewed algorithmic accountability policies:

- *Clear institutional incentives and binding legal frameworks can support consistent and effective implementation of accountability mechanisms, supported by reputational pressure from media coverage and civil society activism:* in contrast to voluntary commitments and non-binding guidance, establishing algorithmic accountability policy through formal legal frameworks ensures that the policy is "on the agenda" of government agencies, and can provide important incentives for implementation, including the potential of judicial or legislative review and oversight.
- *Algorithmic accountability policies need to clearly define the objects of governance as well as establish shared terminologies across government departments:* the lack of standardization and clarity in the definition of the technological systems under governance can turn into an obstacle in interpreting and implementing policy requirements. Adopting broad definitions, particularly in an area where new accountability concerns are constantly being unearthed, can ensure dynamism in the application of these policy mechanisms.
- *Setting the appropriate scope of policy application supports their adoption. Existing approaches for determining scope such as risk-based tiering will need to evolve to prevent under- and over-inclusive application:* agencies should avoid the risk of placing mundane or routine algorithmic processes which do not appear to have significant social impacts under review. Therefore, it is advisable to limit the scope of application of the policy, at least initially, to encourage its adoption within public agencies that may otherwise be hesitant to expend substantial resources on compliance. Possible criteria that may be used to determine whether and when a particular system should be subject to policy scrutiny include: (i) degree of human oversight; (ii) impacts at the individual-level versus group-level impacts; (iii) perceived risk; and (iv) appropriate stage of intervention.
- *Policy mechanisms that focus on transparency must be detailed and audience-appropriate to underpin accountability:* there is the need to design policies for transparency keeping in mind particular audiences and how information can be made usable by them. Also, transparency mechanisms should be designed keeping in mind the potential challenges posed by countervailing policy objectives requiring confidentiality (e.g. privacy, security, intellectual property, risk of systems being gamed), and trade-offs between transparency and other objectives should be negotiated when deciding to use an algorithmic system.
- *Public participation supports policies that meet the needs of affected communities. Policies should prioritize public participation as a core policy goal, supported by appropriate*

resources and formal public engagement strategies: meaningful participation and engagement not only includes the provision of forums for engagement, but also providing educational resources and appropriate time to provide meaningful feedback, and for feedback to be considered and responded to. This is crucial to designing policies in ways that meet the identified needs of affected communities, and in incorporating contextual perspectives that expertise-driven policy objectives may not meet.

- *Policies benefit from institutional coordination across sectors and levels of governance to create consistency in application and leverage diverse expertise*.

e) The European Union

In April 2021, the European Commission proposed a Regulation on Artificial Intelligence ("AI Act"). Even though this draft Act is not focused primarily on governmental usage of AI, the following rules that it proposes for different risk levels of AI are still useful thinking:

- **"Unacceptable risks"** - The following list of usages are prohibited:
    - Manipulative systems likely to cause harm that use subliminal techniques or exploiting vulnerabilities due to age or disability.
    - "Social scoring": AI systems (i) used by or on behalf of public authorities (ii) to generate 'trustworthiness' scores and which (ii) lead to either unjustified or disproportionate treatment of individuals or groups, or detrimental treatment which, while justifiable and proportionate, occurs in an unrelated 'context' from the input data.
    - Some uses of real-time biometric systems (including facial recognition) in publicly accessible spaces by law enforcement, with exceptions.

Scholars Michael Veale and Frederik Zuiderveen Borgesius (2021) see the requirement of "harm" and the limitation to "real-time" as shortcomings of this list. Similarly, Frank Pasquale (2021) considers that this list does not go far enough and suggests that "many forms of nonconsensual A.I.-driven emotion recognition, mental health diagnoses, ethnicity attribution and lie detection should also be banned."

- **"High-risk systems"** - The following list of usages are allowed:
    - Biometric identification and categorization.
    - Management and operation of critical infrastructure.
    - Educational and vocational training.
    - Employment, worker management, and access to self-employment.
    - Access to and enjoyment of essential services and benefits.
    - Law enforcement.
    - Migration, asylum, and border management.
    - Administration of justice and democracy.

However, they should comply with the following requirements and obligations:

- Create a quality management system, which should include a documented risk management system updated throughout the system's lifetime.
- Meet data quality criteria in relation to relevance, representativeness, accuracy, completeness, and application-area specific properties.
- Ensure the accuracy, robustness, and cybersecurity of systems themselves.
- Create technical documentation, readily available for organizations involved in regulation or conformity assessment.
- Build for human oversight, incorporating human-machine interface tools to ensure systems can be effectively overseen by natural persons.

- **"Minimal Risks"** - Usages where the Member States and the Commission merely 'encourage' and 'facilitate' voluntary codes of conduct.

More recently, in the Report on Artificial Intelligence in criminal law and its use by the police and judicial authorities in criminal matters adopted in October 2021, the European Parliament suggested:

1. A moratorium on the deployment of facial recognition systems for law enforcement purposes that have the function of identification, unless strictly used for the purpose of identification of victims of crime, until the technical standards can be considered fully fundamental rights compliant, results derived are non-biased and nondiscriminatory, the legal framework provides strict safeguards against misuse and strict democratic control and oversight, and there is empirical evidence of the necessity and proportionality for the deployment of such technologies.
2. A prohibition of the use of AI by law enforcement authorities to make behavioral predictions on individuals or groups based on historical data and past behavior, group membership, location, or any other such characteristics, thereby attempting to identify people likely to commit a crime.
3. A ban on the use of private facial recognition databases in law enforcement.
4. A ban on AI-enabled mass scale scoring of individuals.

In addition, and in relation to the use of AI in general in judicial and law enforcement contexts, it called for:

- Algorithmic explainability, transparency, traceability, and verification as a necessary part of oversight.
- The adoption of appropriate public procurement processes for AI systems, to ensure their compliance with fundamental rights and applicable legislation, including ensuring that software documentation and algorithms are available and accessible to the competent and supervisory authorities for review.

- Traceability of AI systems and the decision-making process that outlines their functions, defines the capabilities and limitations of the systems, and keeps track of where the defining attributes for a decision originate, through compulsory documentation.
- Compulsory fundamental rights impact assessments to be conducted prior to the implementation or deployment of any AI systems, to assess any potential risks to fundamental rights.
- Periodic mandatory auditing of all AI systems by an independent authority, to test and evaluate algorithmic systems, their context, purpose, accuracy, performance and scale, and, once they are in operation, to detect, investigate, diagnose and rectify any unwanted and adverse effects and to ensure the AI systems are performing as intended.
- Suitable professional training and qualifications to ensure that decision-makers are trained about the potential for bias, as the data sets may be based on discriminatory and prejudiced data.

Frameworks for determining whether uses of ADS or AI-enabled profiling should be prohibited:

One of our recommendations in the main body of the report is that "the state should adopt a framework to evaluate state agency use of ADS technology or use of artificial intelligence-enabled profiling to determine whether or not its use should be prohibited." This section surveys the frameworks and justifications for prohibiting ADS uses in other jurisdictions.

King County's May 2021 analysis of Ordinance 19296 succinctly summarizes the three categories of concerns that have led to this and other prohibitions on facial recognition technology:

- Accuracy of the technology.
- Demographic biases.
- Encroachment on civil liberties.

Santa Cruz' ordinance prohibiting acquisition and use of predictive policing and facial recognition technology highlights several reasons for the ban:

- The propensity for these technologies "to endanger civil rights and civil liberties outweighs these technologies' purported benefits."
- The technologies currently "appear to have the propensity to potentially exacerbate racial injustice."
- The technologies "currently lack the protections needed to adequately safeguard the rights and liberties of all people."

The EU's Draft AI Bill prohibits "systems whose use is considered unacceptable as contravening Union values, for instance by violating fundamental rights." Paragraphs 15-21 of the proposal discuss the rationale for prohibiting certain uses. An excerpt:

> (15) Aside from the many beneficial uses of artificial intelligence, that technology can also be misused and provide novel and powerful tools for manipulative, exploitative and social control practices. Such practices are particularly harmful and should be prohibited because

they contradict Union values of respect for human dignity, freedom, equality, democracy and the rule of law and Union fundamental rights, including the right to non-discrimination, data protection and privacy and the rights of the child….

(17) AI systems providing social scoring of natural persons for general purpose by public authorities or on their behalf may lead to discriminatory outcomes and the exclusion of certain groups. They may violate the right to dignity and non-discrimination and the values of equality and justice.

(18) The use of AI systems for 'real-time' remote biometric identification of natural persons in publicly accessible spaces for the purpose of law enforcement is considered particularly intrusive in the rights and freedoms of the concerned persons, to the extent that it may affect the private life of a large part of the population, evoke a feeling of constant surveillance and indirectly dissuade the exercise of the freedom of assembly and other fundamental rights. In addition, the immediacy of the impact and the limited opportunities for further checks or corrections in relation to the use of such systems operating in 'real-time' carry heightened risks for the rights and freedoms of the persons that are concerned by law enforcement activities.

Glossary

Algorithm: a process or set of rules. Algorithms can be very simple ("to compute sales tax, multiply the price by 9.5%") or extremely complex.

Artificial Intelligence (AI): a general category of computer systems performing tasks that have historically required human intelligence. Annex I of the EU Draft AI Regulation lists specific AI techniques, including machine learning approaches, logic- and knowledge-based approaches, and statistical approaches.

Audit trails: a detailed, chronological record (or set of records) tracking the activity of a system and its users. Audit trails can assist in detecting and diagnosing biases as well as security violations, performance problems, and application defects.

Auditing: a process intended to provide insight into and document the functioning and potential impact of an ADS. Algorithmic Accountability for the Public Sector notes that "regulatory audits are increasingly designed to capture the broader social consequences of a system's use and assess its functioning with respect to an established normative standard, in order to identify potential areas of concern."

Bias: outcomes which are systematically less favorable to individuals within a particular group and where there is no relevant difference between groups that justifies such harms.

Identified or identifiable natural person: a person who can be readily identified, directly or indirectly. This definition applies in the other recommendations as well.

Machine Learning: a form of artificial intelligence in which algorithms are refined (and theoretically improved) based on experience. One common approach to machine learning is to "train" the system based on historical data. Biases in the training data can be reflected in the algorithms and the output of those algorithms.

Monitoring: ongoing analysis of a system to detect inaccuracies and bias.

Natural Persons or Person: a human being (see RCW 46.04.356).

Subpopulation: in statistics, a subgroup of the overall population sharing one or more properties; for example, Washingtonians with disabilities are a subpopulation of Washingtonians.

Testing: providing a system with known inputs and examining the outputs to check for inaccuracy and bias. Testing is a specific technique that can also be useful as part of quality assurance or an algorithmic audit.

Transparency: the ability for people to see and understand a systems' behavior. For ADS, this can include knowing what decision is made and the basis for it, the algorithms that are used to make the decision, whether the system has been audited for biases and the results of the audit, and whether an ADS is used to make or support a specific decision.

Acknowledgments

Thank you to the ADS Workgroup members who contributed their time and expertise to creating this report. Specifically, we want to thank our presenters to help the Workgroup understand the complicated topics being discussed. We also want to thank the advocacy members and subject matter experts for their valuable contributions. Immense gratitude is also due to our state agency members for their time and efforts to this important work. The quality of analysis and civil discourse as the Workgroup navigated the sensitive and complicated areas of ADS and AI impacts was critical to the findings and recommendations in this report. Thank you.