# Mitigating AI Risks in State Government

## December 5, 2023 | 1:00-2:00pm ET

NATIONAL GOVERNORS ASSOCIATION

EPI Center

AAAS

CENTER FOR SCIENTIFIC EVIDENCE IN PUBLIC ISSUES

# Agenda

- **Opening speaker: Alexandra Reeve-Givens, CEO, Center for Democracy and Technology**
  - What are the major categories of risk posed by AI in state governmental applications? How is the federal government approaching risk management?

- **Panel Discussion: Perspectives from State Leaders**
  - **Katy Ruckle**, Chief Privacy Officer, Washington Technology Solutions
  - **Andrew Wheeler**, Director, Office of Regulatory Management, Virginia

- Audience Questions, Moderated Discussion with all Speakers and Panelists

NATIONAL GOVERNORS ASSOCIATION

EPI Center | AAAS
CENTER FOR SCIENTIFIC EVIDENCE IN PUBLIC ISSUES

CENTER FOR DEMOCRACY & TECHNOLOGY

# Advancing Responsible AI: Opportunities for States

December 5, 2023

*Alexandra Reeve Givens*
*President & CEO, Center for Democracy & Technology*

# Center for Democracy & Technology

CDT is a nonprofit, nonpartisan organization based in Washington D.C. and Brussels.

We fight for technology to support public good while protecting against invasive, discriminatory and exploitative uses.

We:

- advocate for sound laws & policies
- advise companies & government on responsible tech use and design.

CENTER FOR DEMOCRACY & TECHNOLOGY

**Agenda:**

1. What types of risks?
2. Elements of trustworthy AI
3. Steps executive agencies can take:
   - ➢ Government's own use of AI
   - ➢ Responding to harmful uses of AI

# The News Cycle About Future AI Harms:

**A.I. poses human extinction risk on par with nuclear war, Sam Altman and other tech leaders warn**

PUBLISHED TUE, MAY 30 2023•11:11 PM EDT | UPDATED WED, MAY 31 2023•8:25 AM EDT

Technology

**Elon Musk and others urge AI pause, citing 'risks to society'**

By **Jyoti Narayan, Krystal Hu, Martin Coulter** and **Supantha Mukherjee**

April 5, 2023 8:22 AM EDT · Updated 7 months ago

Tech Boss Warns of 25% Chance AI Could Destroy Human Civilization

by Roman Perkowski in !!!, Artificial intelligence, Artificial intelligence on 21 October 2023

# The Reality of Current AI Harms:

## Security & Surveillance

- Facial recognition
- Predictive policing
- Risk assessment in bail & sentencing

## Education

- Student activity monitoring
- Remote exam proctoring
- Personalized learning

## Consumer Fraud & Abuse

- Fraud
- Cybersecurity
- Extortion
- Sexual content

## Commercial Data Practices

- Employment
- Housing
- Lending
- Insurance
- Ad Targeting

## Benefits & Public Health

- Eligibility determinations
- Fraud detection
- Medical research & spending

## Information Harms & Elections

- Deepfakes
- Voter suppression
- Targeting & filter bubbles

References:
Blueprint for an AI Bill of Rights (Oct. 2022)(Appendix)
OMB Draft Guidance for Federal Agencies Use of AI (list of presumed rights- and safety-impacting uses)

CENTER FOR DEMOCRACY & TECHNOLOGY

# 1. Example: AI & Public Benefits

TIME

## States' Automated Systems Are Trapping Citizens in Bureaucratic Nightmares With Their Lives on the Line

Michigan Department of Attorney General

State of Michigan Announces Settlement of Civil Rights Class Action Alleging False Accusations of Unemployment Fraud

October 20, 2022

## SSI Recipients' Benefits No Longer Wrongfully Terminated

NYLAG New York Legal Assistance Group

Automating care US news

## What happened when a 'wildly irrational' algorithm made crucial healthcare decisions

## FEDERAL COURT RULES AGAINST IDAHO DEPARTMENT OF HEALTH AND WELFARE IN MEDICAID CLASS ACTION

Ruling Mandates Important Protections for Due Process Rights of Idahoans with Developmental Disabilities

# 2. Other Areas

- Criminal Justice System
- Policing
- Hiring & workplace
- Housing
- Credit


U.S. NEWS

**Facial recognition technology jailed a man for days. His lawsuit joins others from Black plaintiffs**

**Feds Warn Employers Against Discriminatory Hiring Algorithms**

As AI invades the interview process, the DOJ and EEOC have provided guidance to protect people with disabilities from bias.

**U.S. renters fall foul of algorithms in search for a home**

# Elements of Trustworthy AI

## Blueprint for an AI Bill of Rights

1. You Should Be Protected From Unsafe Or Ineffective Systems
2. You Should Not Face Discrimination By Algorithms And Systems Should Be Used And Designed In An Equitable Way
3. You Should Be Protected From Abusive Data Practices Via Built-In Protections And You Should Have Agency Over How Data About You Is Used
4. You Should Know That An Automated System Is Being Used And Understand How And Why It Contributes To Outcomes That Impact You
5. You Should Be Able To Opt Out, Where Appropriate, And Have Access To A Person Who Can Quickly Consider And Remedy Problems You Encounter.

## NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0)

AI Risks & Trustworthiness

- 3.1 Valid & Reliable
- 3.2 Safe
- 3.3. Secure & Resilient
- 3.4 Accountable & Transparent
- 3.5 Explainable & Interpretable
- 3.6 Privacy-Enhanced
- 3.7 Fair, with Harmful Bias Managed

CENTER FOR DEMOCRACY & TECHNOLOGY

# The Role for States: Government Use of AI

## 1. Mandate Risk Management Practices

- ➢ Determine if rights or safety impacting
- ➢ Require minimum practices:
  - Complete AI impact assessment
  - Test performance in real-world context
  - Independently evaluate the AI
  - Conduct ongoing monitoring & threshold for human review
  - Ensure adequate human training
- ➢ Additional minimum practices for rights-impacting uses:
  - Test for equity & nondiscrimination (pre & post deployment)
  - Consult impacted groups
  - Notify impacted individuals at time of encounter
  - Human consideration & remedy; opt-out
- ➢ Comply with due process & APA requirements!

Reference:
OMB Proposed Memorandum for Federal Agency Use of AI (Nov. 1, 2023)

# The Role for States: Government Use of AI

## 2. Require Reporting & Documentation

- ➢ Direct agencies to inventory their uses
- ➢ Issue templates for reporting (internal & public)

## 3. Designate Appropriate Staff; Equip for Success

- ➢ Chief AI officers
- ➢ Provide teams with relevant expertise
- ➢ Guidance; templates; working groups; other support

## 4. Take Specific Steps on Procurement

- ➢ Guidance, templates, staffing support
- ➢ Ensure sufficient control & ownership over data, data improvements, & procured systems
- ➢ Ensure quality control, privacy & security!

References ct'd:


VT AI Inventory (Dec 2022)

CA Executive Order (Sep. 2023)

VA Executive Directive (Sep. 2023)

WA Tech Gen AI Guidelines (Sep. 2023)

# The Role for States: Countering Harmful AI Uses

## 1. Combatting Fakes & Providing Authoritative Information

- Government officials must act to protect their role as trusted sources of civic information.
  - ➢ consistent branding & trust indicators; e.g. .gov domains
  - ➢ proactive messaging & “pre-bunking” false narratives
  - ➢ establish trusted channels for communication



**FBI Warns of AI Deepfake Extortion Scams**

The agency says criminals are turning photos and videos into explicit content to extort victims.

**Deepfake Audio Is a Political Nightmare**

British fact-checkers are racing to debunk a suspicious audio recording of UK opposition leader Keir Starmer.

# The Role for States: Countering Harmful AI Uses

2. **Take AI-driven harms seriously!**

   - Ensure law enforcement is equipped to address consumer fraud, extortion, NCII, election interference
   - Address critical infrastructure & cybersecurity risks
   - Direct housing / civil rights / consumer protection teams to issue guidance & bring cases
   - Any AI funds must require responsible innovation
   - Advance a strategic legislative agenda

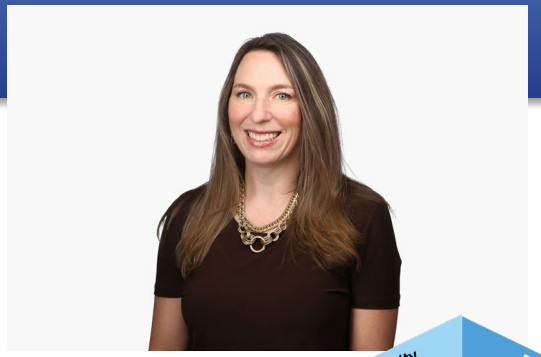
THANK YOU | CDT.ORG

Alexandra Givens
agivens@cdt.org

CENTER FOR DEMOCRACY & TECHNOLOGY

WaTech
Washington Technology Solutions

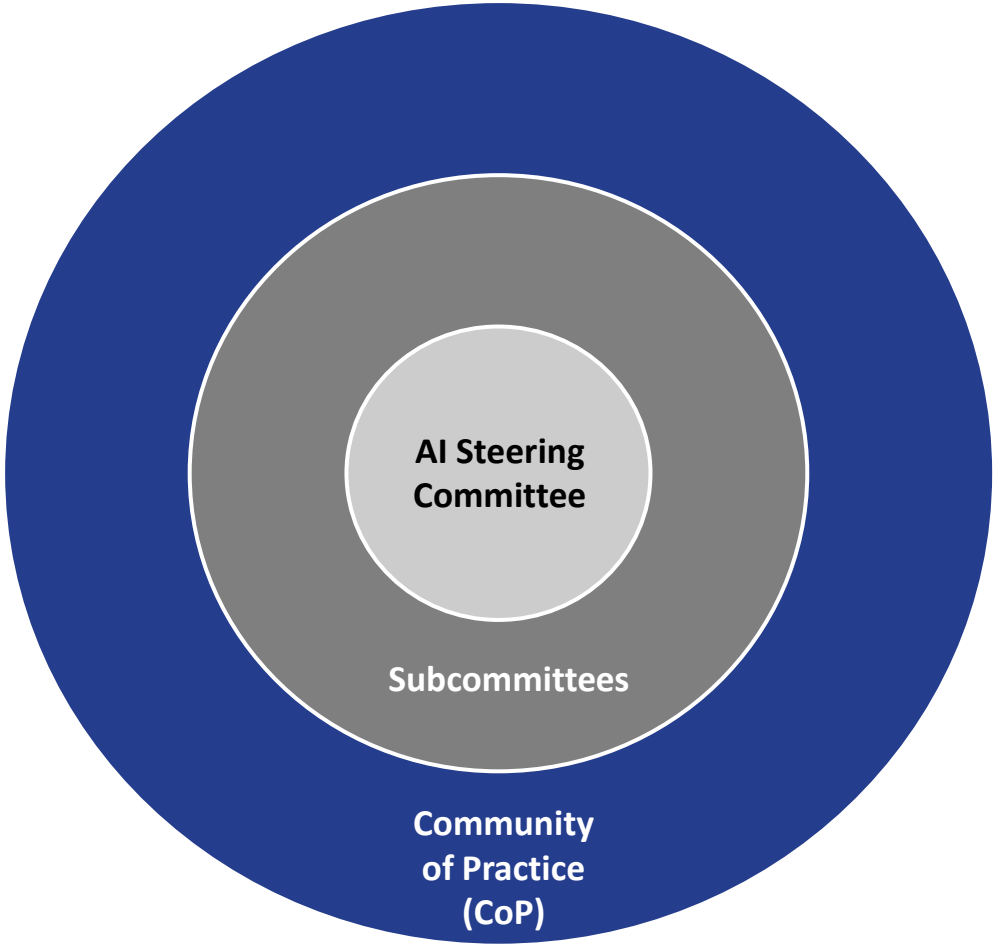
# State Chief Privacy Officer



- Katy Ruckle, State Chief Privacy Officer
- Position created Washington law –
  - Privacy Principles
  - Projects that involve personally identifiable information (PII)
  - Data Protection
- What is CPO role in relationship to AI?
  - Automated Decision Systems Work
  - Generative AI

# AI CoP

- **Governance Structure**
  - Representation from WaTech, State Agency, and Local Government
- **Steering Committee Objectives**
  - Develop a set of **guidelines** and **policies**
  - Identify and document **best practices**
  - Establish a **governance structure** and develop mechanisms for accountability and oversight
  - **Document use cases** and examine potential societal impact
  - **Facilitate collaboration** and knowledge sharing
  - **Promote alignment** of new AI technologies to business and IT strategies



# More data is needed to:

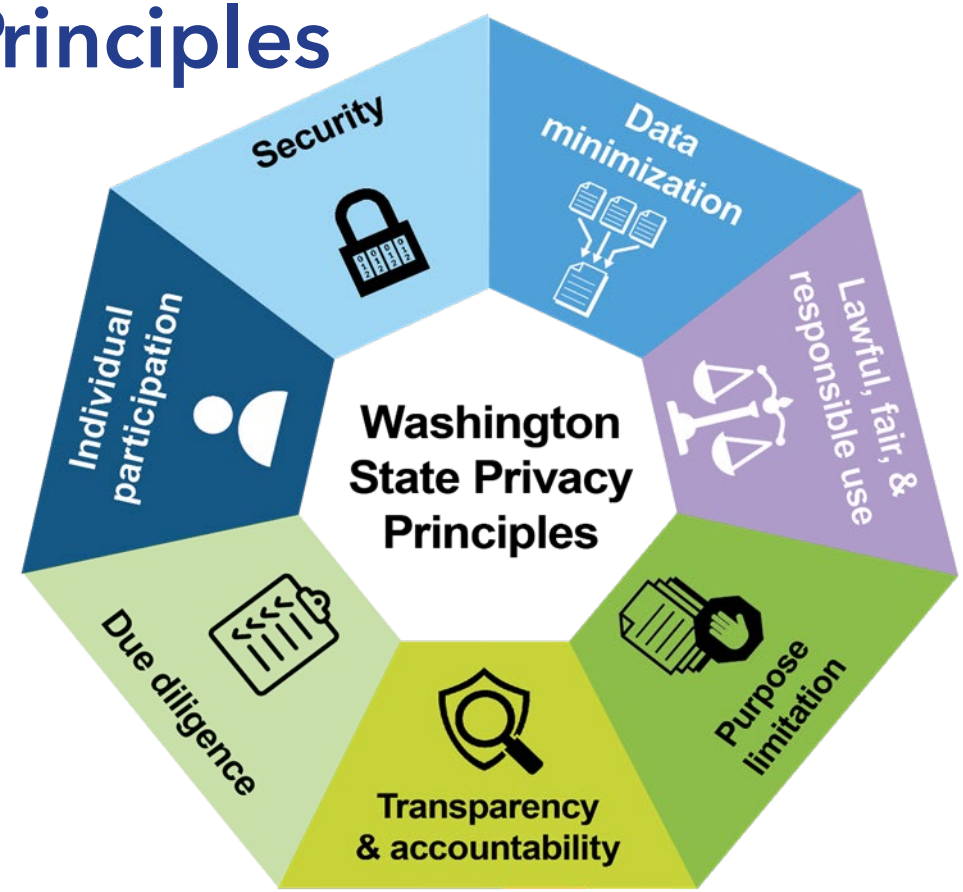
- ✓ Build AI
- ✓ Train AI
- ✓ Maintain AI

# What is the issue with more data from a privacy perspective?

- **Risk of**
  - **Data persistence**
  - **Data repurposing**
  - **Data spillovers**
  - **Data commingling**
  - **Data integrity**

# Washington State Agency Privacy Principles

- Lawful, fair, & responsible use
- Data minimization
- Purpose Limitation
- Transparency & accountability
- Due diligence
- Individual participation
- Security

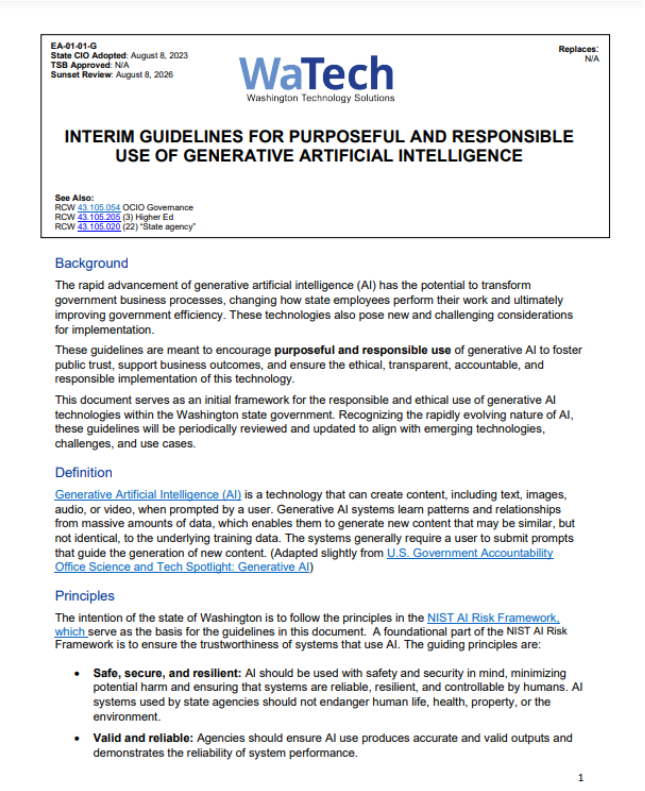



Washington State Agency Privacy Principles

WaTech
Washington Technology Solutions

# https://ocio.wa.gov/policy/generative-ai-guidelines

- Interim Guidelines for Purposeful and Responsible Use of Generative Artificial Intelligence
  - Background
  - Definition
  - Principles
  - Guidelines
  - Generative AI Usage Scenarios and Dos and Don’ts
  - Use Cases
  - Acknowledgments



EA-01-01-G
**State CIO Adopted**: August 8, 2023
**TSB Approved**: N/A
**Sunset Review**: August 8, 2026

**Replaces:** N/A

WaTech
Washington Technology Solutions

**INTERIM GUIDELINES FOR PURPOSEFUL AND RESPONSIBLE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE**

**See Also:**
RCW 43.105.054 OCIO Governance
RCW 43.105.205 (3) Higher Ed
RCW 43.105.020 (22) "State agency"

**Background**

The rapid advancement of generative artificial intelligence (AI) has the potential to transform government business processes, changing how state employees perform their work and ultimately improving government efficiency. These technologies also pose new and challenging considerations for implementation.

These guidelines are meant to encourage **purposeful and responsible use** of generative AI to foster public trust, support business outcomes, and ensure the ethical, transparent, accountable, and responsible implementation of this technology.

This document serves as an initial framework for the responsible and ethical use of generative AI technologies within the Washington state government. Recognizing the rapidly evolving nature of AI, these guidelines will be periodically reviewed and updated to align with emerging technologies, challenges, and use cases.

**Definition**

Generative Artificial Intelligence (AI) is a technology that can create content, including text, images, audio, or video, when prompted by a user. Generative AI systems learn patterns and relationships from massive amounts of data, which enables them to generate new content that may be similar, but not identical, to the underlying training data. The systems generally require a user to submit prompts that guide the generation of new content. (Adapted slightly from U.S. Government Accountability Office Science and Tech Spotlight: Generative AI)

**Principles**

The intention of the state of Washington is to follow the principles in the NIST AI Risk Framework, which serve as the basis for the guidelines in this document. A foundational part of the NIST AI Risk Framework is to ensure the trustworthiness of systems that use AI. The guiding principles are:

- **Safe, secure, and resilient:** AI should be used with safety and security in mind, minimizing potential harm and ensuring that systems are reliable, resilient, and controllable by humans. AI systems used by state agencies should not endanger human life, health, property, or the environment.
- **Valid and reliable:** Agencies should ensure AI use produces accurate and valid outputs and demonstrates the reliability of system performance.

1

WaTech
Washington Technology Solutions

# Guiding Principles

- Safe, secure, and resilient
- Valid and reliable
- Fairness, inclusion, and non-discrimination
- Privacy and data protection
- Transparency and auditability
- Accountability and responsibility
- Explainable and interpretable
- Public purpose and social benefit

# Guidelines for Generative AI Use

# Guidelines

## • Fact-checking, Bias Reduction, and Review

- All content generated by AI should be reviewed and fact-checked, especially if used in public communication or decision-making.
- State personnel generating content with AI systems should verify that the content does not contain inaccurate or outdated information and potentially harmful or offensive material.
- Given that AI systems may reflect biases in their training data or processing algorithms, state personnel should also review and edit AI-generated content to reduce potential biases.
- When consuming AI-generated content, be mindful of the potential biases and inaccuracies that may be present.

# Guidelines

## • Disclosure and Attribution

- AI-generated content used in official state capacity should be clearly labeled as such, and details of its review and editing process (how the material was reviewed, edited, and by whom) should be provided. This allows for transparent authorship and responsible content evaluation.
- State personnel should conduct due diligence to ensure no copyrighted material is published without appropriate attribution or the acquisition of necessary rights. This includes content generated by AI systems, which could inadvertently infringe upon existing copyrights.

# Guidelines

## • Sensitive or Confidential Data

- Agencies are strongly advised not to integrate, enter, or otherwise incorporate any non-public data (non-Category 1 data) or information into publicly accessible generative AI systems (e.g., ChatGPT).
- If non-public data is involved, agencies should not acquire generative AI services, enter into service agreements with generative AI vendors, or use open-source AI generative technology unless they have undergone a Security Design Review and received prior written authorization from the relevant authority, which may include a data sharing contract.
- Contact your agency's Privacy and Security Officers to provide further guidance.

# State Ethics law – Confidential Information

- RCW 42.52.050

(3) No state officer or state employee may disclose ***confidential information*** to any ***person*** not entitled or authorized to receive the information.

- Definitions (RCW 42.52.010):

(5) "Confidential information" means (a) specific information, rather than generalized knowledge, that is not available to the general public on request or (b) information made confidential by law.

(15) "Person" means any individual, partnership, association, corporation, firm, institution, or other entity, whether or not operated for profit.

# Generative AI Usage Scenarios Do's and Don'ts

# ✓ Do's (best practices) and x Don’ts (things to avoid)

➢ **Rewrite documents in plain language for better accessibility and understandability.**

✓ **Do** specify the reading level in the prompt, use readability apps to ensure the text is easily understandable and matches the intended reading level, and review the rewritten documents for biases and inaccuracies.

➢ **Condense longer documents and summarize text.**

✓ **Do** read the entire document independently and review the summary for biases and inaccuracies.

x **Don’t** include sensitive or confidential information in the prompt

# Do's (best practices) and Don’ts (things to avoid)

## ➢ Draft Documents

✓ **Do** edit and review the document, label the content appropriately, and remember that you and the state of Washington are responsible and accountable for the impact and consequences of the generated content.

X **Don’t** include sensitive or confidential information in the prompt or use generative AI to draft communication materials on sensitive topics that require a human touch.

# Do's (best practices) and Don’ts (things to avoid)

## ➢ Aid in Coding

✓ **Do** understand what the code is doing before deploying it in a production environment, understand the use of libraries and dependencies, and develop familiarity with vulnerabilities and other security considerations associated with the code.

**X Don’t** include sensitive or confidential information (including passwords, keys, proprietary information, etc.) in the prompt and code

# Do's (best practices) and Don’ts (things to avoid)

- **Aid in generating image, audio, and video content for more effective communication**

✓ **Do** review generated content for biases and inaccuracies and engage with your communication department before using AI-generated audiovisual content for public consumption.

**x Don’t** include sensitive or confidential information in the prompt.

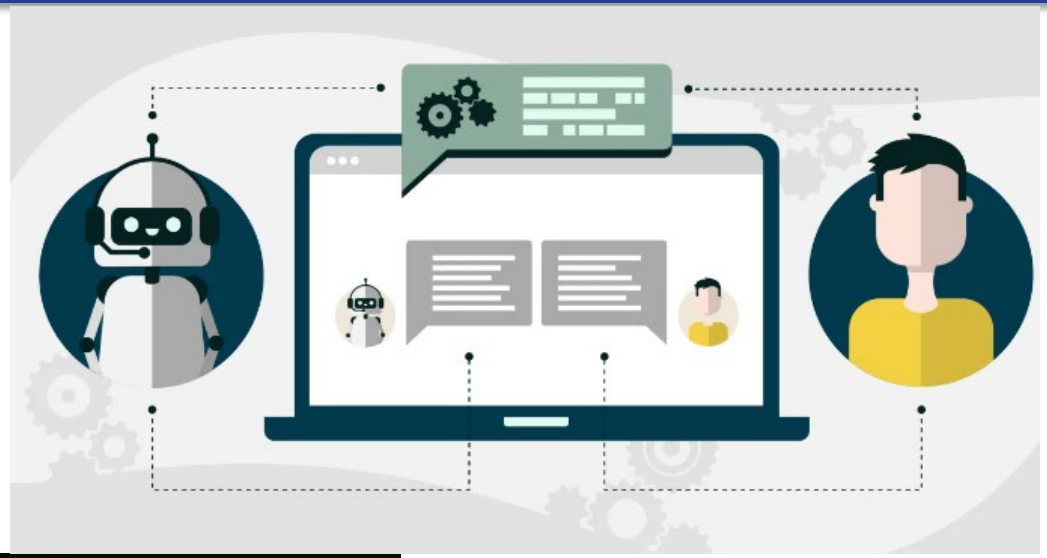
# Do's (best practices) and Don’ts (things to avoid)

## ➢ Automate responses to frequently asked questions from residents (example: chatbots)

✓ **Do** implement robust measures to protect resident data.

**X Don’t** use generative AI as a substitute for human interaction or assume it will perfectly understand residents’ queries. Provide mechanisms for residents to easily escalate their concerns or seek human assistance if the AI system cannot address their needs effectively.

# Use Cases



# Other data and privacy considerations for Generative AI?

- Where did the training data come from?
- Was the training data legally obtained?
- Data being used as a proxy for something else?

# Artificial Intelligence Regulation in Washington

- **SSB 5116 (2021)** - Establishing guidelines for government procurement and use of **automated decision systems** in order to protect consumers, improve transparency, and create more market predictability.

POLICY

## Lawmakers Move to Ban Discriminatory Tech in Washington State

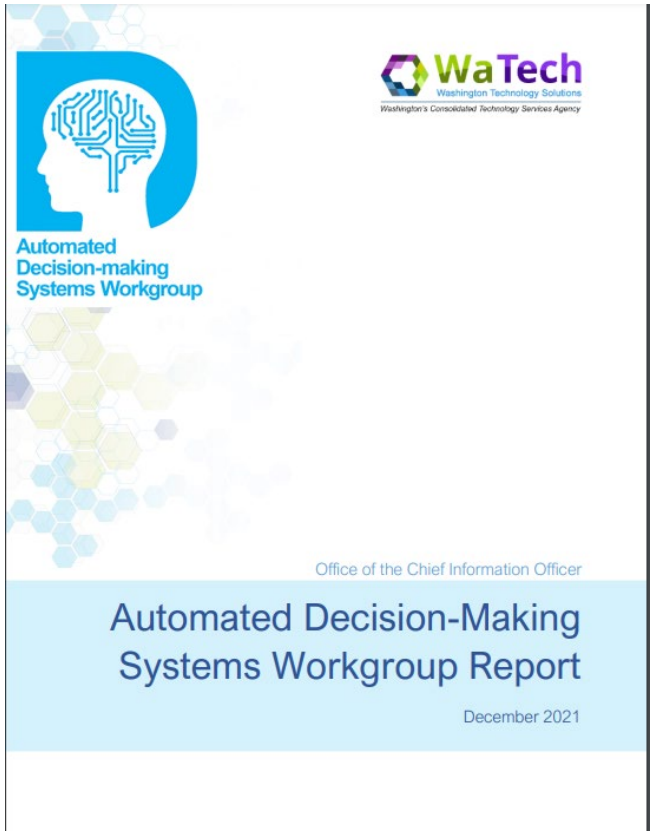
In response to reports detailing AI tech's disproportionate impact on communities of color, Washington State Sen. Bob Hasegawa introduced a bill to ban AI tech and regulate automated decision systems.

February 23, 2021 • Katya Maruri

# 2021 Report & Recommendations

- #1 Prioritization of Resources
- #2 Procurement
- #3 Evaluation of Existing Systems
- #4 Transparency
- #5 Determination on Whether to Use System
- #6 Ongoing Monitoring or Auditing
- #7 Training in Risk of Automation Bias

- 2021 ADS Workgroup Report



WaTech
Washington Technology Solutions

# Questions?

privacy@watech.wa.gov

# AI Resource List

- Please see webinar **Chatbox for a link** to the list

- Includes:
  - Federal level activities
  - State activities: Executive Branch and Legislative
  - Local activities
  - Technical assistance tools

NATIONAL GOVERNORS ASSOCIATION

EPI Center | AAAS
CENTER FOR SCIENTIFIC EVIDENCE IN PUBLIC ISSUES

# Contacts


- Kate Stoll, AAAS EPI Center: kstoll@aaas.org
- Sally Rood, NGA Center for Best Practices: srood@nga.org
- Ryan Martin, NGA Center: rmartin@nga.org


NATIONAL GOVERNORS ASSOCIATION

EPI Center | AAAS
CENTER FOR SCIENTIFIC EVIDENCE IN PUBLIC ISSUES