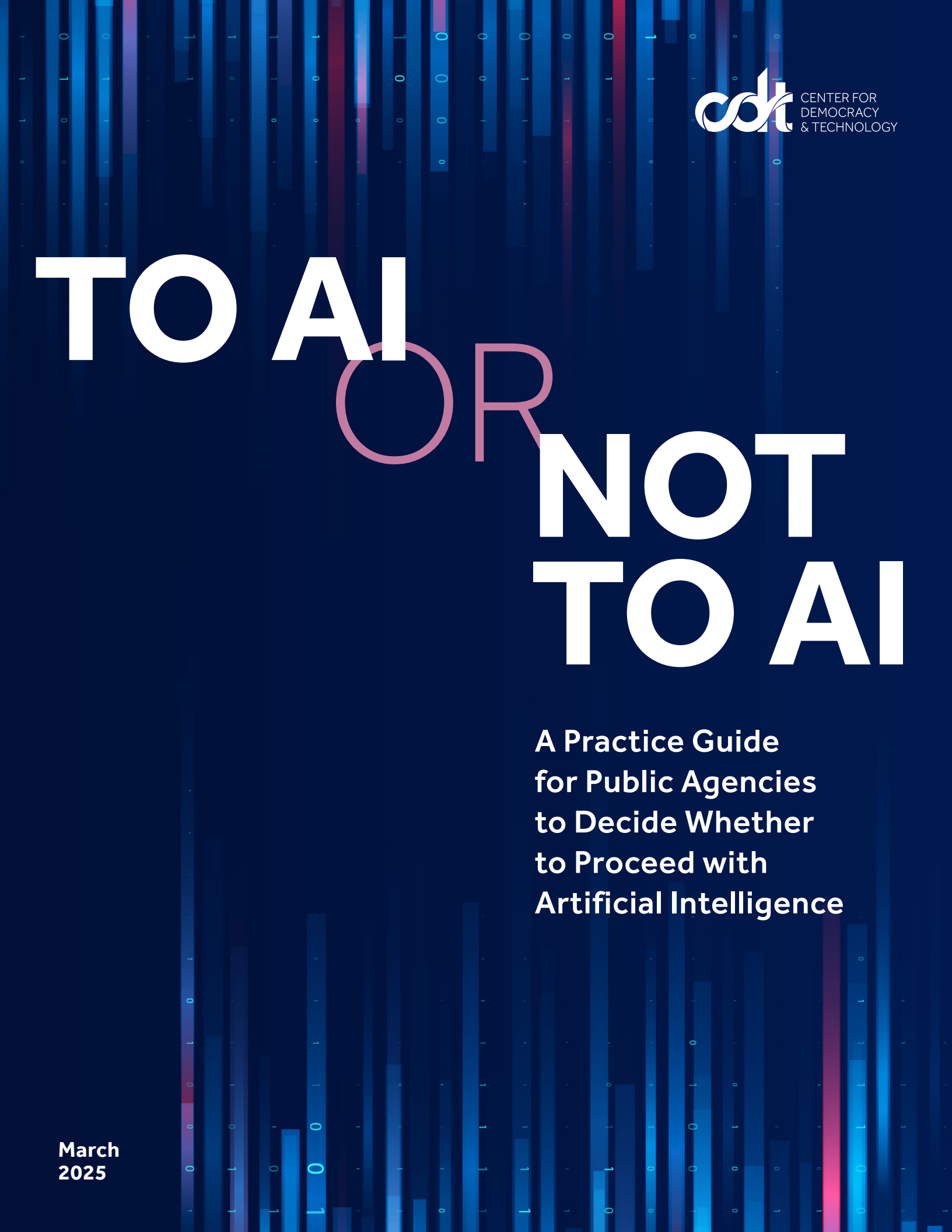

TO AI OR NOT TO AI

A Practice Guide for Public Agencies to Decide Whether to Proceed with Artificial Intelligence

March 2025

AUTHOR:

Sahana Srinivasan
Elizabeth Laird

The Center for Democracy & Technology (CDT) is the leading nonpartisan, nonprofit organization fighting to advance civil rights and civil liberties in the digital age. We shape technology policy, governance, and design with a focus on equity and democratic values. Established in 1994, CDT has been a trusted advocate for digital rights since the earliest days of the internet. The organization is headquartered in Washington, D.C. and has a Europe Office in Brussels, Belgium.

As governments expand their use of technology and data, it is critical that they do so in ways that affirm individual privacy, respect civil rights, foster inclusive participatory systems, promote transparent and accountable oversight, and advance just social structures within the broader community. CDT's Equity in Civic Technology Project furthers these goals by providing balanced advocacy that promotes the responsible use of data and technology while protecting the privacy and civil rights of individuals. We engage with these issues from both technical and policyminded perspectives, creating solutions-oriented policy resources and actionable technical guidance.

Endnotes in this report include original links as well as links archived and shortened by the Perma.cc service. The Perma.cc links also contain information on the date of retrieval and archive.

Table of Contents

Executive Summary

Public agencies have significant incentives to adopt artificial intelligence (AI) in their delivery of services and benefits, particularly amid recent advancements in generative AI.[1] In fact, public agencies have already been using AI for years in use cases ranging from chatbots that help constituents navigate agency websites to fraud detection in benefit applications.[2] Agencies' resource constraints,[3] as well as their desire to innovate, increase efficiency, and improve the quality of their services,[4] all make AI and the potential benefits it often offers — automation of repetitive tasks, analysis of large swaths of data, and more[5] — an attractive area to invest in.

However, using AI to solve any problem or for any other agency use case should not be a foregone conclusion. There are limitations both to AI's capabilities generally and to it being a logical fit for a given situation.[6] Thus, agencies should engage in an explicit decision-making process *before* developing or procuring AI systems to determine whether AI is a viable option to solve a given problem and a stronger solution than non-AI alternatives. The agency should then repeatedly reevaluate its decision-making throughout the AI development lifecycle if it decides initially to proceed with an AI system. Vetting the use of AI is critical because inappropriate use of AI in government service and benefit delivery can undermine individuals' rights and safety and waste resources.[7]

Despite the emergence of new frameworks, guidance, and recommendations to support the overall responsible use of AI by public agencies, there is a dearth of guidance on how to decide whether AI should be used in the first place, including how to compare it to other solutions and how to document and communicate that decision-making process to the public. This brief seeks to address this gap by proposing a four-step framework that public administrators can use to help them determine whether to proceed with an AI system for a particular use case:

- **Identify priority problems for the public agency and its constituents:** Agencies should identify and analyze specific problems they or their constituents face in service or benefit delivery to ensure that any new innovations are targeted to the most pressing needs. Agencies can identify problems and pain points in their service and benefit delivery through mechanisms such as existing agency data, news reports, and constituent engagement and feedback. Agencies should then vet the severity of their problem and set specific and measurable goals and baselines for what they hope their eventual solution accomplishes.

- **Brainstorm potential solutions to priority problems:** Agencies should identify a slate of solution options for their problem. These options may include AI systems but should also consider non-AI and nontechnological alternatives. Desk research, landscape analyses, consultation with other government agencies, and preliminary conversations with vendors can help agencies ensure that they have identified all options at their disposal before potentially focusing on AI. This report will detail preliminary options for solutions to common agency problems, including AI-based and non-AI options.

- **Evaluate whether AI could be a viable solution before comparing alternatives:** Agencies need to evaluate each potential solution on a set of criteria tailored to that solution before deciding on one with which to proceed. This guidance presents an *AI Fit Assessment:* four criteria that agencies can use to evaluate any solution that involves an AI-based system. Agencies can use this resulting analysis to decide whether proceeding with an AI-based solution is viable. Agencies should adopt rubrics, no-go criteria, green flags, or other signals to determine how their evaluations of solutions on these four criteria correspond to proceeding with or forgoing a solution. They should also reevaluate the *AI Fit Assessment,* their analysis of alternatives, and their decision to use AI throughout the development process, even if they initially decide to proceed with an AI-based solution. The criteria of the *AI Fit Assessment* are the following:
  - ***Evidence base:*** the level of evidence demonstrating a particular AI system's capabilities, effectiveness, and appropriateness, specific to the use case and including evidence of its strengths over alternative solutions.
  - ***Data quality:*** the availability and quality of data, from either the vendor or the agency, used to power the solution as well as the ethics of using that data.
  - ***Organizational readiness:*** the agency's level of preparedness to adopt and monitor AI, including its infrastructure, resources, buy-in, and technical talent.
  - ***Risk assessments:*** the results of risk and/or impact assessments and any risk mitigation plans.

  The results of the *AI Fit Assessment* will provide agencies with an analysis of an AI solution, which they can then weigh against separate analyses of non-AI alternatives to finally determine which solution to initially proceed with. While non-AI solutions can be evaluated using the *AI Fit Assessment,* not all of the questions will apply, and additional analysis may be needed.

- **Document and communicate agency decision-making on AI uses to the public:** For at least all use cases in which they decide to proceed with an AI-based solution, agencies should document the analysis from the preceding three action steps — including their analysis of AI-based solutions, analysis of non-AI alternative solution options, and comparison of the options — and communicate these insights to the public. Communicating the rationale behind their AI use cases to the public helps agencies build constituents' trust in both the agency itself and in any AI systems constituents interact with.[8] For the sake of transparency and to help others navigate similar use cases, agencies can also consider documenting situations in which they decided against AI.

Because this brief refers to any form of AI system when discussing AI, including algorithms that predict outcomes or classify data, the guidance can be used when considering whether to proceed with any type of AI use case.

Most importantly, these action steps should assist public administrators in making informed decisions about whether the promises of AI can be realized in improving agencies' delivery of services and benefits while still protecting individuals, particularly individuals' privacy, safety, and civil rights. This decision-making process is especially critical to navigate responsibly when public agencies are considering moderate- or high-risk AI uses that affect constituents' lives and could potentially affect safety or human rights.

Why Active Decision-Making on Whether to Use AI Matters

Potential Benefits of AI in Public Service Delivery

The pressures and incentives that public agencies face to use artificial intelligence (AI) in their delivery of services and benefits primarily fall into three categories. First, the efficiency and performance benefits that AI systems can offer can help agencies meet existing internal goals to improve their service and benefit delivery, specifically by easing resource constraints or logistical inefficiencies.[9] Many agencies also generally desire to innovate and modernize, which could encourage the use of unnecessary — or at least insufficiently vetted — AI.[10]

Second, increased hype around AI — bolstered significantly by some true successes in generative AI since 2022 — and significant marketing have increased the pressure on all types of organizations to adopt it.[11] Ensuring that any new AI systems public agencies develop or procure can actually do what they advertise for a particular use case is critical for these agencies.[12]

Third, public agencies must adhere to the increasing number of federal, state, and local policy mandates requiring them to modernize, cut down on inefficiency, and improve their services. The 2021 Executive Order on Transforming Federal Customer Experience and Service Delivery to Rebuild Trust in Government declared that federal agencies must design and deliver services "with a focus on the actual experience of the people whom [they are] meant to serve" — in other words, that agencies must prioritize the quality of their customer service experience, not just the quality of the services themselves.[13] Simultaneously, one of the most common use cases of AI among federal agencies is generative AI-powered chatbots that act in customer service-oriented roles, assisting constituents in navigating agency websites or answering questions.[14] Most of the 13 states (and the District of Columbia) that have released executive orders (EOs) on AI in the past two years have also mentioned the imperative for state agencies to take the fullest advantage of AI advancements by proactively identifying new AI opportunities and specifically wielding AI to improve their delivery of services and benefits.[15]

Potential Risks of AI in Public Service Delivery

Amid and before these pressures, federal, state, and local agencies have used AI in myriad ways, some of which demonstrate how the use of AI systems can go wrong. AI has inherent qualities that make it different from other digital technologies: its computing power, ability to "learn" from large quantities of data, the fact that it often relies on categorizing data,[16] and the tendency of people using AI to defer to its judgment in decision-making processes even when they are supposed to have the final call, for example.[17] Sometimes AI goes wrong simply because a model is not good enough at its task.[18] Any of these qualities may make the integration of AI a poor fit for a situation and can lead to government deployment of AI that worsens rather than improves agencies' ability to deliver services and benefits.

For example, New York City released a generative AI-powered chatbot in October 2023 that was meant to issue advice and answer questions on starting and running a small business. The chatbot ended up sometimes giving out incorrect information, including calling illegal landlord practices such as discriminating against applicants on the basis of protected characteristics permissible.[19] The chatbot's website now includes a disclaimer that the technology is still being tested and sometimes doles out inaccurate information.[20] Generative AI chatbots' inherent tendency to hallucinate, or generate misleading or incorrect information framed as fact, is well documented.[21]

At the height of the COVID-19 pandemic, the California Employment Development Department procured an AI system to detect fraudulent public benefit claims. The system ended up having an extremely high false positive error rate, meaning that real people with valid claims were often identified as submitting fraudulent claims and denied benefits. The AI system led to the invalidation of 1.1 million claims, more than half of which were legitimate.[22]

In another example, several counties and states procured biased or ineffective AI systems meant to predict instances of child abuse to help them decide whether to further investigate flagged families.[23] In some instances, these systems led to families with lower economic status being investigated at what some argue were unfairly, disproportionately high rates.[24]

Policy and Legal Requirements for Active Decision-Making About AI

States are trying to guide their agencies' deployment of AI by creating review processes for potential AI uses or requiring agencies to proactively consider alternatives to AI before moving forward with an AI system. For example, California's guidance on public sector generative AI procurement specifies that preprocurement analysis must include research on the problem the agency is trying to solve with AI and market research to develop a slate of options for solutions to that problem.[25] Virginia's AI Policy Standard specifies that to use AI agencies must determine that it is the best

solution to a problem among multiple considered alternatives. Virginia also requires that agencies analyze the cost and benefits of AI's alternatives.[26] As of the publication of this brief, 13 states in addition to the District of Columbia have issued EOs focused on AI. Most of those EOs emphasize that state agencies should identify and prioritize their most promising, effective AI use cases.

Cost-Benefit Analysis of AI Use in Public Service Delivery

Even if no policy or legal mandates require an agency to justify its use of AI, public administrators should still evaluate whether AI is better than alternatives throughout the AI lifecycle. First and foremost, this diligence protects constituents from unnecessary risks of AI. AI's benefits can sometimes be worth its risks when weighed together; however, AI with its considerable potential harms should not be used when it cannot solve the problem it was designed for in the first place.

This concept is closely related to the proportionality principle espoused by international guidelines on AI. The proportionality principle states that any solution, including any AI-based one, should be tailored and right-sized to the problem being solved for and not affect constituents unnecessarily.[27] In other words, AI should be used only if it is both effective and necessary. Similar questioning about the use and necessity of sharing personal data is also a best practice in the data privacy field, a space in which the risk and benefit must constantly be balanced as they are with AI.[28]

Cost-benefit analysis in general is a well-established government practice: Federal agencies have been required since the 1970s to conduct this type of analysis for any rulemaking expected to have significant economic impacts.[29] Similarly, before they are designed, procured, or deployed, potential AI use cases should be analyzed comprehensively, including but not limited to their potential risks and harms. The idea that decision-making on whether to proceed with AI should be an explicit step within its development lifecycle also has precedence. The fields of sociology and psychology break the innovation process into three stages: design and planning of the innovation, a decision on whether to proceed, and then actual implementation.[30] Much of the existing responsible AI guidance discusses risk mitigation practices that organizations can engage in during the design and planning phase or the implementation phase, but there is less support for what agencies should do in the middle step that is focused on deciding whether to proceed.

Related Resources

Despite the number of imperatives to justify the use of AI that now exist, and the plethora of guidance on responsible AI in general, most agencies focus on how to mitigate the risks of AI rather than how to decide for or against it. For example, the National Institute of Standards and Technology AI Risk Management Framework (NIST AI RMF) extensively outlines risk management practices and lists acceptable AI use cases from a risk perspective but not from perspectives of cost, benefit, fit, or other factors that should be used to decide whether to proceed with an AI use case.[31] Most AI ethics frameworks and guidelines focus on high-level principles and not how to operationalize those principles into decisions on whether to use AI. They also often omit factors outside of ethical principles, such as AI's evidence-based fit for a situation.[32]

AI risk assessments are frameworks that typically focus on identifying the potential negative impacts of an AI system and risk mitigation practices that the developer or deployer can then integrate into the system. The recommendations and framework for deciding whether to use AI in the first place that are introduced in the next section include consideration of a risk assessment's results but also incorporate factors such as organizational readiness, the evidence base for success in this situation, and more. Risk assessments, model evaluations, audits, and other concrete ways of analyzing AI's effects should contribute to making decisions on whether to use AI but are insufficient on their own to make informed decisions about whether and how to use AI in the delivery of public services and benefits.

Some academic literature emphasizes the importance of vetting AI's functionality, organizational governance factors, organizational readiness factors, and other societal and ethical factors before using AI.[33] This guidance in part draws upon and cites these works, aiming not only to consolidate prior work on this topic but also to provide an actionable framework for public agencies to act on these factors.

How to Determine Whether to Use AI

Given the hype and excitement that surrounds AI, public administrators need to exercise caution to separate fact from fiction. Although AI has the potential to improve the delivery of public services, accounting for its limitations is equally important before deciding to move forward with this technology. These limitations may indicate that an agency has to mitigate the risks of an AI system, or they may indicate that the agency should not proceed with an AI-based solution at all.

There are general rules of thumb to consider when deciding whether AI is appropriate to use for a given problem. Humans tend to be better, and AI weaker, at interventions and decisions in complex environments that cannot easily be modeled.[34] In fact, some uses of AI show significant evidence of overall ineffectiveness. For example, AI systems tend to be inherently ineffective at predicting social outcomes — including recidivism, child abuse risk, and job performance — and at decision-making that hinges on predicting constituents' behavior, such as denying or approving loan claims based on predicting defaults.[35] AI, on the other hand, tends to excel in automating repetitive tasks; personalizing or customizing processes; and handling certain writing tasks: writing or editing documents and emails, reformatting content, and writing and sorting through code.[36]

Though these rules of thumb can be used as initial heuristics in informing whether to proceed with AI, this report's guidance — particularly the *AI Fit Assessment* (described in more detail on page 21) — delineates a more comprehensive process and criteria that agencies can leverage to decide whether an AI-based solution is appropriate and ultimately whether to use AI to solve a particular problem. Determining whether AI should be used in the first place is arguably the most important decision point in the AI lifecycle. It is too often assumed that AI can help and is the best fit to achieve a goal.[37] However, it is critical to question that assumption and repeatedly ask and answer whether AI is a viable approach and to analyze whether it is a stronger solution than non-AI alternatives.

To determine whether to proceed with AI, public agencies should take the following steps:

1. **Identify priority problems for the public agency and its constituents:** Identify and analyze specific problems public administrators and their constituents face in service or benefit delivery to ensure that any new innovations are targeted to the most pressing needs.

2. **Brainstorm potential solutions to priority problems:** Identify a slate of solution options for the problems that the public agency has prioritized. These options may include AI but should also consider non-AI and nontechnological alternatives.

3. **Evaluate whether AI could be a viable solution before comparing alternatives:** Make an explicit decision on whether to initially proceed with an AI solution by evaluating the merits and costs of all solution options across a variety of criteria. This decision requires agencies to evaluate AI-based solutions on the four criteria of the *AI Fit Assessment*, evaluate non-AI alternatives, and then compare their full slate of solution options to decide on the best one to initially proceed with.

4. **Document and communicate agency decision-making on AI uses to the public:** Document the analyzed problems, brainstormed solutions, and merits and costs of solution options across criteria (i.e., the analysis of the preceding three action steps); provide justification for proceeding with an AI-driven solution (if that is the decision); and communicate these insights to the public in a digestible manner.

Although public administrators should make active decisions and use the *AI Fit Assessment* throughout the AI lifecycle, as detailed in the callout box on page 13, this guidance focuses on how agencies can decide during early planning phases whether to proceed with an AI system in the first place. The appendix provides additional details on how the criteria and guiding questions of the *AI Fit Assessment* can be adapted later in the AI lifecycle.

When to Determine Whether to Use AI

Agencies should analyze problems and brainstorm solution options (the first two recommendations introduced starting on page 14) as early in the process as possible, but they may revisit these steps later on.

Using the *AI Fit Assessment* to determine whether AI is a viable solution — the key action involved in the third recommendation of this guidance — is not a one-time decision. Even if agencies initially find AI to be viable and better than alternatives and decide to move forward with using AI, the development and deployment phases of AI systems have important inflection points during which agencies should affirmatively decide whether to continue to use AI.[38] The resulting analysis from the *AI Fit Assessment* will necessarily evolve as agencies procure or build out solutions, test, and monitor them.[39] Indefinite use of AI to solve a particular problem should not be deemed a foregone conclusion. Additionally, agencies should not get trapped in a sunk cost fallacy: Even if they have already expended resources on an AI system that they later determine, based on any of the *AI Fit Assessment* criteria, is no longer appropriate to continue, they should make an affirmative decision to end its use.

An example of when the recommended action steps of this guidance, particularly the use of the *AI Fit Assessment* to determine if an AI-based solution is viable, could occur is shown below in Figure 1:

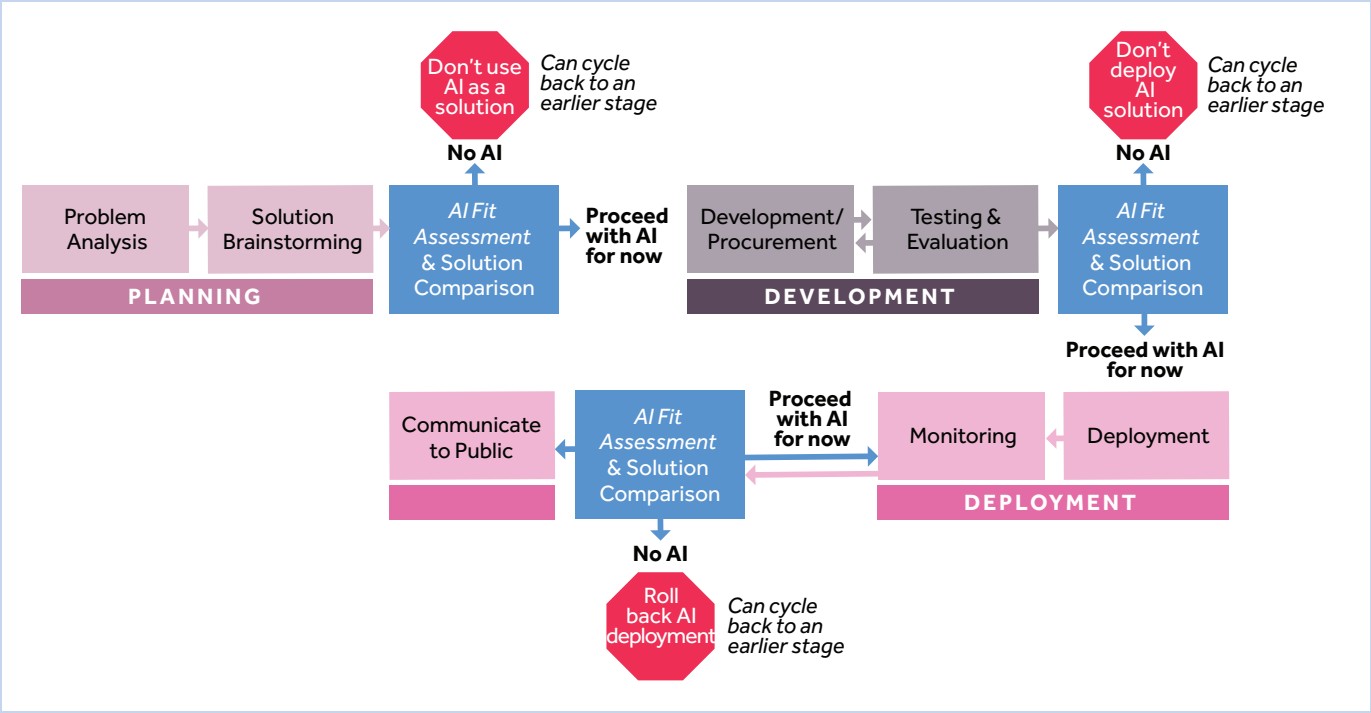

Figure 1: This figure presents a simplified view of an AI development and procurement lifecycle and highlights the points at which public administrators should make active decisions about whether to use AI.[40]

RECOMMENDATION 1

Identify Priority Problems for the Public Agency and Its Constituents

The agency should identify a problem, validate that it is significant enough to merit investing in a robust solution,[41] and establish measurable goals, which can be used to vet potential solutions and evaluate whether a chosen solution has led to progress toward the agency's original goals.[42]

In addition to mandated goals and initiatives, which are often established through external directives such as EOs and legislation, public agencies should identify problems that must be addressed to enable them to achieve their overall mission and serve the needs of their constituents.[43] In identifying priority problems, public agencies would benefit from including the perspectives of internal and external stakeholders and soliciting this input through multiple methods such as interviews and data analysis.[44] Consulting with constituents, particularly marginalized groups or groups for whom the problem is the most significant, through a variety of mechanisms and providing needed support for their participation can help the agency identify the most pressing issues its constituents face and more precisely pinpoint problems to solve.[45]

Once a slate of problems has been selected, public agencies should also assess their urgency.[46] The agency should validate that the problem it has identified is impeding its delivery of services and benefits to constituents in some significant way to confirm that solving the problem would be impactful and a priority among other identified problems.[47]

Finally, the agency should craft goals related to the problem it is trying to solve. This work includes documenting the status quo, establishing baseline metrics related to the agency's goal, and identifying the agency's ideal target for those metrics once a solution has been implemented so that it can measure the performance of the eventual solution.[48] The agency's goals should be as precise as possible and include quantifications wherever possible. Setting clear goals upfront enables agencies to decide proactively if their solution options — including any AI-based ones — are capable of meeting those goals and then to evaluate, once a solution has been selected and developed, if it actually does.

Which Internal Stakeholders Should Be Involved in Internal Governance of AI?

The recommendations included in this guidance are best achieved when they are incorporated into existing data and technology governance processes and are implemented by people with a variety of skill sets, including technical, legal, policy, and programmatic expertise.

EXISTING DATA AND TECHNOLOGY GOVERNANCE PROCESSES

Some agencies may already have processes in place that guide their development of technology generally. This report's guidance should not override those processes but should instead be integrated into them as appropriate. The stakeholders involved in any existing data and technology governance processes can use this report's criteria to help make their decisions. Figure 1 on page 13 of this report can be mapped to the agency's existing data and technology governance processes to identify precisely where the *AI Fit Assessment,* the comparison of AI and non-AI solution options, and the other recommendations of this guidance should be incorporated into existing structures.

MULTIDISCIPLINARY SUBJECT MATTER EXPERTISE

The recommendations in this guidance and the required decision-making should be implemented by a variety of stakeholders who represent a range of expertise, including experts in social science.[49] Stakeholders involved in the decision-making process could include an agency's chief AI officer, chief data officer, chief technology officer, general counsel, and policy and/or program directors. The research needed to implement the recommendations could be undertaken either by these individuals or by other agency employees, with the final decision-making authority then lying with these stakeholders. This guidance should always be implemented in consultation with a variety of external stakeholders and constituents.[50] Agencies should engage with their communities and consult other agencies, vendors, and civil society and civil rights groups as needed, particularly for AI systems they have identified as high risk or that have been deemed internally or externally as rights impacting or safety impacting.[51]

RECOMMENDATION 2

Brainstorm Potential Solutions to Priority Problems

Once an agency has identified priority problems, it should conduct research to identify potential solutions. Potential solutions could include maintaining the status quo, as well as implementing a range of options from nontechnical, people-centered processes to those that rely on technology such as non-AI software, generative AI, or other AI-driven solutions.[52] Alternatives to common AI use cases in the delivery of public services and benefits are listed in the call-out box titled "Starter List of AI Use Cases and Alternatives" on page 17.

Key methodologies for this stage include landscape analyses to identify which solutions have been used to address similar prior problems in practice or theoretically; market research on potential solutions that have been deployed; consultation with other agencies to see how they have addressed similar goals or problems; experimentation with solutions or consultation with vendors; community engagement; and internal stakeholder conversations, particularly with the program teams that work on the area most closely related to the agency's problem and have the most contextual knowledge related to any solutions.[53]

Starter List of AI Use Cases and Alternatives

This is a *starter* list of potential public agency AI use cases, informed by existing case studies and literature demonstrating some success and efficacy.[54] This list is meant to help agencies begin brainstorming solution options for particular goals, but agencies should still research their individual use cases and the types of AI systems they are considering in the manner outlined by the third recommendation of this guidance, the *AI Fit Assessment*. This list does not contain all examples of public sector AI use or all alternatives to AI systems, but it includes several common ones.

GOAL	**Improve customer service by providing accurate information about public services faster and more conveniently to constituents.[55]**	
SPECIFIC USE CASES	Enhance website usability.[56]	
SOME POTENTIAL SOLUTIONS	**AI BASED** 1. Develop or integrate generative AI into website chatbots to do things such as:[57] a. Understand user questions and map them to prewritten responses; b. Filter out chat requests that can be solved without human customer service support;[58] and c. Respond directly to user questions.[59] 2. Use natural language processing techniques (machine learning algorithms that can analyze language) to analyze website information and constituent questions to do things such as:[60] a. Identify outdated or incorrect information on the website; and b. Identify common questions for which the agency can develop prewritten answers.	**ALTERNATIVES TO AI** 1. Enhance existing website functionalities such as: a. Menu-based customer service chatbots; and b. Better FAQ pages. 2. Enhance website content pages to show up better in search responses or to better answer constituent questions.

GOAL	**Improve the quality and efficiency of constituents' service and benefit experiences through personalization.**	
SPECIFIC USE CASES	Customize and streamline forms based on analysis of constituent feedback.	
SOME POTENTIAL SOLUTIONS	**AI BASED** 1. Use natural language processing or other machine learning techniques to customize forms based on constituents' earlier responses.[61] 2. Use natural language processing to synthesize constituents' responses and feedback for better personalized experiences in the future.[62]	**ALTERNATIVES TO AI** 1. Use rule-based personalization based on predefined categories (e.g., customers manually answer categorization questions to get a personalized form).

GOAL	**Improve agency efficiency and reduce the cost of services by automating certain operational or administrative processes.[63]**	
SPECIFIC USE CASES	Increase the speed of analyzing or creating documents, datasets, or programming code.	
SOME POTENTIAL SOLUTIONS	**AI BASED** 1. Use generative AI or other natural language processing techniques to fill out, search, or summarize documents used in internal processes.[64] 2. Use algorithms to analyze and draw new insights from large datasets more efficiently than manual review or to automate data entry and processing.[65] 3. Use generative AI to assist in code generation for data analysis or other tasks.[66]	**ALTERNATIVES TO AI** 1. Analyze data using non-AI statistical techniques. 2. Use logic or rule-based data and process automation to reduce the number of steps in human analysis or document creation. 3. Procure software platforms that make manual analysis work easier or more efficient.

GOAL	**Reduce fraud and waste by more effectively and efficiently flagging potential fraud.[67]**	
SPECIFIC USE CASES	Improve ability to detect or flag fraud or anomalies.	
SOME POTENTIAL SOLUTIONS	**AI BASED** 1. Use one of a variety of machine learning techniques to do the following:[68] a. Detect fraudulent documentation;[69] b. Flag anomalous financial, billing, or claim patterns;[70] and c. Detect fraudulent checks before they are cashed.[71]	**ALTERNATIVES TO AI** 1. Use logic or rule-based detection of anomalous data points to flag potential fraud. 2. Develop software platforms that make human review of claims for potential fraud easier or more efficient. 3. Focus on fraud prevention and education instead of detection after the fact. 4. Improve cybersecurity practices, such as two-factor authentication to prevent fraudulent misuses of someone else's identity or detection of duplicate or foreign IP addresses to flag potential foreign attacks or fraudulent duplicate claims.[72]

RECOMMENDATION 3

Evaluate Whether AI Could Be a Viable Solution Before Comparing Alternatives

After identifying problems and brainstorming solutions, agencies should use the *AI Fit Assessment* to decide whether any AI-based systems they are considering are viable solutions to their identified problem. They should then analyze their other, non-AI solution options and compare all solutions to make an explicit decision on whether or not to proceed with an AI-based system. The *AI Fit Assessment* includes four criteria to support agencies' decisions. The key questions needed to address each of these criteria are described later in this section, and a full list of guiding questions for each criterion is presented in detail in the appendix. Agencies can use some of the *AI Fit Assessment*'s criteria and questions to evaluate their non-AI solution options as well, but some questions are relevant only to AI-based systems. Agencies can evaluate the costs and benefits of their non-AI alternatives in the manner they have previously used to evaluate technology options, if applicable.

This guidance includes red flags for each criterion of the *AI Fit Assessment* to highlight some responses that should halt agency work on AI systems for a given use case, at least until further research is completed. However, agencies should build upon this guidance's list of red flags with their own and develop a rubric to guide how their answers to these criteria questions will correspond to decisions on whether or not to proceed with AI-based solutions.

If an agency is procuring its AI system from a vendor, it can ask the *AI Fit Assessment*'s questions to the vendor as part of the procurement process or during preprocurement analysis and calls. Vendor answers should be held to the same scrutiny as answers for internal systems, and if the vendor cannot or chooses not to answer (for instance, if it feels an answer would reveal proprietary information), that response should be considered a red flag.

The Use of AI Is Not All or Nothing

Because of the many forms that AI can take, there is a secondary layer to the question of whether agencies should proceed with AI — namely, they must also decide what *kind* of AI to design, procure, or develop; *what role* it will play within a system; and *how widely* to implement it.[73]

KIND OF AI

A variety of types of models and algorithms are considered AI, including ones that predict outcomes, detect patterns, analyze language, and more. Agencies may be evaluating which types of algorithms or models to use in addition to whether to use AI in the first place. This guidance can be used to help agencies compare different AI systems to each other in addition to comparing AI systems to non-AI alternatives.

ROLE OF AI

The slate of solution options an agency may be considering to solve a given problem could also include different ways or levels of integrating AI into the agency's delivery of services and benefits. For example, in attempting to improve customer service on their website with a new chatbot, agencies might consider a chatbot entirely based on large language models (e.g., the model parses and generates responses to user questions) or one only partially based on a model (e.g., the model parses questions but maps them to fixed, prewritten answers).

The *AI Fit Assessment* and the other recommendations in this guidance can be used to compare a slate of solution options that include either one or multiple AI systems integrated into agency processes in different ways. However, this report does not enumerate all the forms AI can take and how extensive its implementation should be.

HOW WIDELY TO IMPLEMENT AI

Additionally, agencies may seek to develop pilots of AI systems before deciding whether to procure or develop a system that they deploy to constituents.[74] The *AI Fit Assessment*'s criteria and questions can also help agencies decide which use cases are best suited to AI pilots or whether to proceed with existing pilots.

The Four Criteria of the *AI Fit Assessment* to Evaluate Whether AI Is a Viable Solution

EVIDENCE BASE

To inform decision-making about whether to proceed with an AI system and ensure that AI is actually the best solution for a problem, it is important to understand the level of evidence demonstrating a particular AI system's capabilities, including its strengths over alternative solutions.[75] The evidence should be as specific as possible to the AI system being considered and the context and use case to which it will be applied.[76] The types of evidence available and necessary for this criterion will vary depending on the phase of the AI lifecycle (described in Figure 1 on page 13). In the initial planning phase, gathering this evidence will require primarily market and vendor research, review of academic research, engagement with impacted communities,[77] and consultation with experts to study the effectiveness and appropriateness of AI in solving similar problems, both in theory and at other organizations.[78] Existing evidence schema can provide a helpful starting point for agencies to quantify and evaluate their levels of supporting evidence.[79]

GUIDING QUESTIONS

- What evidence is there that this AI-based system can meet the agency's specific goals,[80] such as addressing constituent needs,[81] saving money, or otherwise creating efficiency?[82]
- What does existing evidence indicate about whether this AI-based system is more or less effective than alternative solutions in this context, including non-AI and nontechnical solutions?[83]
- What evidence is there that using AI in this way will maintain the agency's quality in delivering services and benefits?

RED FLAGS

For this criterion, the red flags could indicate that AI is not a viable solution in this use case or that the agency needs to do more research before proceeding with an AI-based solution.

- **Lack of foundational information and efficacy evidence:** Neither the agency nor the vendor (if technology is being procured) is able to answer a substantial portion of the questions in this framework for the applicable use case and type of AI system under consideration. The effectiveness of a potential AI solution is supported by little evidence, or the evidence weighs against the effectiveness or appropriateness of a potential AI solution in this context. For example, the agency is trying to predict social outcomes, a task at which AI systems are ineffective.

- **Lack of demonstrable benefits:** The agency is unable to highlight serious benefits of AI over alternatives. Or AI is considered generally helpful but not necessarily effective in achieving the agency's specific goals.

- **Lack of community engagement in the evidence-gathering process for use cases that affect the constituent community:** No engagement with the community has occurred to identify their most pressing issues, which is critical to understanding whether there is evidence that a given solution can meet constituents' needs. This gap could result in a situation in which constituents' and/or employees' needs and goals do not align with the agency's and/or with the strengths of the AI-based solution. This red flag indicates the agency should at least do more research with the community before proceeding with the AI-based solution.

DATA QUALITY

AI systems will only be as good as the data on which they are trained; therefore, it is critical that public agencies vet whether they or their vendor (or potentially both, as needed) already have, or can practically and ethically acquire and use, sufficient data of a high enough quality to train an AI system[84] since AI systems require massive datasets with rich information that is highly relevant to the task.[85] This work involves assessing the accuracy, completeness, representativeness, and other characteristics of the data, as well as investigating and strengthening the processes that generated and will continue to generate data for the AI system.[86] Public agencies must also ensure that the data they are leveraging and the manner in which they plan to use it comply with relevant legal restrictions, including domain-specific ones.[87]

GUIDING QUESTIONS

- Is using the data needed for this AI-based solution responsible? Is the data unbiased, complete, representative, and legally and ethically able to be used for these purposes (e.g., if it was collected for a different original purpose)?[88]

- How well can key information relevant to the agency's context (i.e., the decision, process, or goal that the agency is hoping to use AI to help with) be captured in data that the organization already has or can get?[89] If the agency is considering a pretrained model, how relevant is the data the vendor already used to the agency's context?

- Is the agency's data reliable? Will the agency or vendor consistently get new data to update the model if needed,[90] or is the historical data still relevant and useful?[91]

RED FLAGS

- **Underlying bias:** The data has representation bias (i.e., some groups are disproportionately over- or under-represented in the data the model is trained on) or has not been explored for bias.[92]

- **Legal or ethical limitations on secondary uses:** The data that will be or has already been used to develop the AI-based system was originally collected for a different purpose, and there are either legal restrictions on this secondary use or ethical concerns if the current scope of use diverges from what the original data subject consented to or was aware of.[93]

- **Proxy data used:** Certain data features are being used as a proxy for real-world data the agency does not have and cannot get, in a manner that is risky (e.g., engagement with public services is being used as a proxy for family stability in the child welfare context).[94]

ORGANIZATIONAL READINESS

In addition to the quality of information available for an AI system, agencies should consider other organizational readiness dimensions that will inform whether the agency is prepared to adopt AI. This work includes vetting the AI system's resourcing and organizational and technical infrastructure. It also includes considering whether the agency is prepared to bear the upfront and ongoing costs of AI development and risk mitigation, whether it has appropriate talent in place to responsibly use AI and a full understanding of the use case at hand, whether it has the flexibility to experiment and iterate, and more.[95] Another consideration is the level of organizational buy-in, particularly from senior leadership, technical talent, and the constituents and employees who would be interacting with or affected by any deployed AI system.[96] Regardless of whether the AI solution is being developed in house or procured, the success of an AI system depends in part on the agency environment being ready for it.[97]

GUIDING QUESTIONS

- Does the agency have a chief AI officer (or other senior leader) who is overseeing the agency's AI strategy and cross-departmental issues such as procurement, training, bias mitigation, and public transparency?
- Has the agency established a process through which AI-powered solutions will be reviewed, iterated on, and/or built (for procurement and/or internal development)?[98]
- Does the agency have appropriate talent to develop, manage, and/or supervise the development of AI, including team members with sociotechnical expertise and experts in trust and safety?[99]
- Is the agency's information technology infrastructure (e.g., its data storage capabilities, networking capabilities, and computing power) ready for the procurement or development of AI?[100]
- Does the agency have long-term resources available to manage the AI system, such as funding and prioritization in agency goals and strategic planning?
- Is the agency culturally prepared, both internally and externally, for the integration of AI?[101] How do the agency's users, constituents, and employees feel about AI being used in this way?[102] Do they seem prepared for and bought into its integration?

RED FLAGS[103]

- **Lack of governance processes:** The agency has not developed or procured AI yet and has not set up governance processes or internally strategized on its AI policies, process, and priorities. The agency has not developed a process to vet, evaluate, and monitor AI.

- **Lack of AI talent:** The agency does not have internal talent to either develop AI internally or question and monitor vendor AI software.

- **Lack of infrastructural support for AI:** The agency does not have the technical infrastructure to support deploying or building an AI system. For example, it may be trying to develop AI-based systems internally but lack scalable computing power, or it may lack the networking capabilities to process and transport large amounts of data quickly to either train or operate AI-based systems.[104]

- **Lack of internal or constituent buy-in for AI:** The agency has major internal concerns about the necessity or benefit of AI. These concerns could come from leadership, constituents, AI talent, or the employees who will be working with any developed AI system (e.g., in evaluating benefit claims).[105]

! RISK ASSESSMENTS

A risk assessment that includes identifying and investigating AI harms, such as civil rights issues, data privacy, security, and other potential impacts, should be conducted before AI is decided on and especially before it is deployed in a real-world context and with a significant number of constituents. Risk assessments should collectively cover as much as possible the potential risks incurred from the AI algorithm itself and the overall system it is being deployed in.[106] Both immediate and longer-term risks should be considered. Once risk assessments have been conducted, organizations must determine the risk mitigations that are feasible and that they are willing and able to implement. They also must either decide that the remaining risks of a system are within their level of tolerance or stop proceeding with an AI system in this use case.[107] Agencies should return to and reevaluate their risk assessments throughout the AI development lifecycle, especially once a vendor or system is chosen and when it is tested. The agency should begin conducting a risk assessment as early as possible, including before deciding whether to proceed with an AI system, but may be able to answer certain questions only later in the procurement or development process.

GUIDING QUESTIONS[108]

- What is the agency's acceptable level of risk and impact (i.e., the potential consequences it is willing to bear and/or the levels of any metrics that quantify the risk of harm) that AI-powered systems could have on key areas such as civil rights, physical and psychological safety, economic opportunity, educational access, the environment, individuals' or the agency's reputations, and more?[109] Risk can be quantified as a function of the probability of harm occurring and the degree of severity of the harm. Levels of risk tolerance should be decided internally by the agency.[110]

- If the agency already has vendor proposals or a specific AI system in mind, has the agency conducted a preliminary risk assessment of the aforementioned risks and others included in frameworks such as the NIST AI RMF?[111]
  - What are the data privacy, security, and environmental impacts of the system?[112]
  - Are the level and likelihood of these risks within the agency's accepted level of risk tolerance, taking into account any planned mitigations? What are the agency's plans to mitigate the risks it has identified and ensure that they remain within its level of tolerance?

- How do the anticipated risks, harms, and other impacts of this AI solution compare to those of any alternatives?

- What are the agency's tolerance levels for any quantifiable safety, fairness, ethics, or other similar criteria or metrics, and does research suggest that achieving those tolerances is possible with the kind of AI-based systems the agency is considering?[113]

RED FLAGS

- **Lack of consensus on risk threshold:** The agency has not yet determined its acceptable level of risk and impact of AI systems.

- **Lack of research on risk likelihoods:** The agency has not researched the likelihood of an AI solution creating or exacerbating risk in the areas outlined previously, such as civil rights, physical safety, economic opportunity, and more, and in risk assessment frameworks such as the NIST AI RMF.[114]

- **Unacceptable risks:** The AI solution would create risks that run afoul of existing laws such as privacy laws and categorically should be avoided regardless of the agency's risk tolerance threshold.

- **Lack of risk mitigation plans:** The agency does not have risk mitigation plans in place, or there is no evidence that the risks the agency has identified can be mitigated at all. An extreme example is that even state-of-the-art models cannot mitigate a risk of harm that the agency has identified as above its risk tolerance threshold.

Once the agency has evaluated one or more AI-based solution options on these four criteria to verify that the options are viable, the agency should compare its analysis to that of non-AI alternatives to decide the best solution with which to initially proceed.

RECOMMENDATION 4

Document and Communicate Agency Decision-Making on AI Uses to the Public

Individuals who regularly interact with public agencies as well as the general public benefit from better understanding whether and how the government is using AI.[115] Transparency is also a critical tool to hold public agencies accountable for their decision-making.[116] However, given the technical and detail-oriented nature of the *AI Fit Assessment*, public agencies need to strike the right balance in communicating their thought processes in ways and in venues that are easily understood by nontechnical audiences.

Generally, public agencies can effectively communicate their decision-making process about whether to use AI to the public if they:

1. **Clearly state the specific problem(s) the agency is trying to solve and the alternatives it considered:** Document answers to the work done in this guidance's first two recommendations around identifying problems and brainstorming solutions at a high level. At minimum, include the agency's identified problems, goals, and slate of solution options considered, including any baselines or metrics the agency will use to evaluate the performance and success of solutions later on.[117]

2. **Communicate how the agency applied the *AI Fit Assessment* to determine whether to proceed with AI:** Document high-level summary results for each criterion of the *AI Fit Assessment*, either answering each guiding question in turn or synthesizing key takeaways across the guiding questions to explain how the agency decided whether a given AI solution was viable. Important types of information to include are:
   - High-level summaries of the insight for each criterion;
   - Summaries of any evaluations or assessment results that were used to address a criterion; and
   - The main non-AI alternatives that the AI system was compared to, a high-level summary of their key characteristics, and an explanation of why the final solution was chosen.

3. **Communicate summaries of the agency's internal thought processes and cost-benefit analyses:** Document the overall decision made on whether or not to proceed with AI (e.g., the "no" or "yes, for now" answer) as well as a high-level summary of how the results of each criterion and the agency's comparison of its AI-based solution options and alternatives contributed to this decision.[118]

4. **Communicate the methodologies and processes the agency used:** Include a high-level summary of the methodologies used in acting on each recommendation and each criterion of the *AI Fit Assessment* (e.g., community engagement, literature review, consultations with internal stakeholders). The agency should also provide a high-level summary of the overall AI development and review process used and the points when key decision steps on AI and other steps of the AI development process occurred (i.e., a diagram similar to this report's Figure 1). This transparency will help the public concretely understand the processes the agency went through in addition to its findings.[119]

5. **Provide clear and concise explanations:** Ensure that all agency explanations of its decision-making are written in plain language and include citations to any third-party evidence.[120] Include high-level summaries and link to more detailed internal documentation on the agency's AI development processes whenever possible.

6. **Format all public reporting accessibly:** Date all public reporting on the agency's AI uses, including all updates, to ensure that constituents have the fullest possible understanding of which AI use cases are active and the current justifications for and analyses of them.[121] Agencies should also ensure that their reporting is machine readable.[122]

Conclusion

As public agencies consider opportunities to adopt AI into their delivery of services and benefits, sorting through the hype and marketing around AI is essential to critically analyze if it will genuinely benefit the agency and the agency's constituents. Public agencies must make active, explicit decisions, repeatedly and throughout the AI development lifecycle, on whether to proceed with AI. These decisions must consider the evidence base in favor of the solution, the available data quality, the agency's readiness for AI, and the potential risks of the AI system. Agencies' analysis of the benefits and drawbacks of AI solutions should also always be compared against that of alternatives and communicated transparently to the public. Increasingly, states and the federal government are mandating this kind of solution analysis, particularly that agencies compare AI solutions to their alternatives before deploying an AI system.

Beginning this work upfront in the AI development lifecycle saves agency resources by lowering the risk of the agency having to shut down an AI solution after investing significantly in it. Most importantly, making the use of AI an active, considered decision rather than a foregone conclusion protects individuals from AI harms, especially from unnecessary harms caused by ineffective or ill-fitting AI systems, ensuring that agencies can realize the benefits of AI while truly supporting their constituents and their mission.

Appendix: Detailed Guiding Questions for the *AI Fit Assessment* Criteria

QUESTIONS TO ASK BEFORE DEVELOPMENT OR PROCUREMENT OF AI, BY CRITERION

Agencies may have to modify these questions or ask them earlier or later, depending on each agency's existing processes for developing or procuring technology including AI. Agencies may also have to wait on some of these questions until after they have talked with vendors or sent out a request for proposals. If agencies encounter a question they cannot yet answer, they should flag it to return to at a later point in the AI development lifecycle. And even if they have already answered these questions, agencies should circle back to them throughout the AI development lifecycle as they obtain more information about and access to the AI system, including after evaluation and testing. An excellent resource for agencies looking to dive even deeper into some of these questions is *The* Situate AI *Guidebook*.

EVIDENCE BASE[123]

QUESTIONS TO ASK

What evidence is there that this AI-based system can meet the agency's specific goals,[124] such as addressing constituent needs,[125] saving money, or otherwise creating efficiency?[126]

- Is there evidence of AI meeting goals similar to the one the agency has identified (e.g., efficiency) in specific real-world contexts resembling the one in which the system will be deployed (e.g., in customer service)?[127]
- Is there evidence of the AI system meeting its advertised purpose and intended use cases?[128] Is there evidence of this AI system failing?
- Can this task be generalized and/or standardized enough for an AI-based system to use?[129] Or does it seem like highly individualized or contextualized attention from humans is needed?[130]
- What about the status quo is not working?
- Is AI a logical fit for this goal, use case, and context, based on the box titled "Starter List of AI Use Cases and Alternatives" on page 17?
- Is the use of AI proportional here? In other words, is developing AI and undergoing risk mitigation for AI worth the goals the agency has outlined?[131] Is the use of AI narrowly targeted to the agency's goals, or is it also being considered for other reasons?[132] How will the agency ensure that the AI system's use stays narrowly targeted since new risks, mitigations, costs, and benefits must be considered for separate contexts and use cases?
- Does the appropriate use of this AI system, or AI in this context generally, depend on human oversight?[133]

- Have user research, interviews, and other forms of community engagement and/or internal research validated that the problem the agency is solving for exists and is significant enough to require an involved solution?[134]
- How much will this solution benefit the recipient of the agency's services/benefits over alternative solutions, if at all?[135] Or is this AI targeted to internal processes only?
  - What trade-offs has the agency made between constituents' needs and benefits and agency goals, if at all?[136]

What does existing evidence indicate about whether this AI-based system is more or less effective than alternative solutions in this context, including non-AI and nontechnical solutions?[137]

- What are the alternatives to AI, and what (if any) specific benefits over them does AI provide? What goals might AI meet that other solutions do not?
- What evidence is there for AI being more effective than those alternatives, including the current solution?[138]

What evidence is there that using AI in this way will maintain the agency's quality in delivering services and benefits?

- What are the key indicators of performance or metrics that the agency wants to maintain while simultaneously addressing its problem or solving a particular goal?
- What is the agency's level of tolerance for those performance metrics changing due to any new AI solution? What is the process for the agency to decide that the change in metrics or the overall impact of the AI system is too harmful and that the agency will retire it?[139]
- Does the literature review or other research and experimentation give the agency reason to believe those metrics — and the overall quality of the agency's service and benefit delivery — will not be negatively affected by an AI solution?

METHODOLOGIES

- Prior literature, efficacy studies, vendor evaluations, and/or the agency's own preliminary evaluations or experiments
- Literature review, including both vendor-provided information, such as model cards,[140] and third-party evidence, that is as specific to the problem or use case as possible
- Community engagement, such as user studies, surveys, or interviews demonstrating that an AI system is likely to meet the agency's goals and constituents' needs as expected[141]
- Consultation with internal AI and software development experts
- Consultation with whichever employees typically handle the problem or process being addressed and thus have the most contextual knowledge

DATA QUALITY

QUESTIONS TO ASK

Is using the data needed for this AI-based solution responsible? Is the data unbiased, complete, representative, and legally and ethically able to be used for these purposes (e.g., if it was collected for a different original purpose)?[142]

- Do any legal restrictions prevent the agency from using this data for this purpose?[143] Or do any legal restrictions prevent the vendor from having access to and/or processing the data?
- What biases are possible in the data (e.g., selection bias, reporting bias)?[144]
- Has the agency or the vendor explored the data to understand its makeup and potential skews or biases in it? Is the data representative of the population at hand as populations being underrepresented in training datasets can lead to unfair algorithms?[145]
- What techniques are being used to explore and validate the data?[146]

How well can key information relevant to the agency's context (i.e., the decision, process, or goal that the agency is hoping to use AI to help with) be captured in data that the organization already has or can get?[147] If the agency is considering a pretrained model, how relevant is the data the vendor already used to the agency's context?

- Is the data noisy, or does it reflect the real world well?[148]
- Does the agency have concrete evidence that this data is relevant to its goals or outcomes?
- Does either the agency or the vendor have historical data available and in a large quantity, or can one of them acquire it ethically?[149] Is there reason to believe the historical data is still relevant to current and future real-world contexts?[150] Is the effort needed to acquire the data worth the expected benefit of hitting the agency's already-defined goal metrics?

Is the agency's data reliable? Will the agency or vendor consistently get new data to update the model if needed,[151] or is the historical data still relevant and useful?[152]

- How was the data collected, or how will it be collected? Is that process reliable and comprehensive, and will it continue to be that way?[153]
- Are the data sources in the agency's control and transparent? Or are they outputs from an opaque system and/or third party?[154]
- Has the agency vetted the correctness and precision of the data (if not in its entirety, at least sampled at random)?

METHODOLOGIES

- Exploratory data analysis, exploration, and visualization
- Review of internal documentation and data
- Consultation with any internal stakeholders who intake, document, or manage relevant data sources

ORGANIZATIONAL READINESS

QUESTIONS TO ASK

Does the agency have a chief AI officer (or other senior leader) who is overseeing the agency's AI strategy and cross-departmental issues such as procurement, training, bias mitigation, and public transparency?

Has the agency established a process through which AI-powered solutions will be reviewed, iterated on, and/or built (for procurement and/or internal development)?[155]

- Has the agency decided who will be involved in which steps and decisions in this process?
  - What specific tools, templates, or assessments do the stakeholders who will be involved in this process have to help them?
- Does the agency have buy-in and accountability from senior leadership?
- Has the agency talked to developers and the employees who will have to interact with the potential AI system in the deployment about how they feel about this use case? Do they have any objections to this being a reasonable use case for AI after the proposed AI use case has been explained to them?[156]

Does the agency have appropriate talent to develop, manage, and/or supervise the development of AI, including team members with sociotechnical expertise and experts in trust and safety?[157]

- Does the agency have talent that has developed, assessed, and deployed AI before (or worked with vendors to do so)?
- Has the agency's talent been educated on the responsible AI processes, both conceptually and how to operationalize them? Does the agency have talent that reports feeling confident in their ability to procure or develop responsible, safe, ethical AI?[158]
- Are employees and constituents ready to interact with AI? Do they have training on or familiarity with how to do so already, or will the agency be providing it?[159] Are they comfortable with AI?
- Does the agency have AI talent that fully understands the use case, the stakeholders who will interact with it and their needs, and the nuances of the system and real-world environment it is being developed for?

Is the agency's information technology infrastructure (e.g., its data storage capabilities, networking capabilities, and computing power) ready for the procurement or development of AI?[160]

- What technical infrastructure does the agency already have in place, and how secure is it? Does the agency have development processes set up for model development or just regular software development?[161]
- How much computing power does the agency or potential vendors have access to, and is that sufficient to develop the envisioned system? Can the agency store the large amount of data required in some AI solutions?[162]
- Does the agency have robust data privacy and security practices in place?
- Does the agency have tools to test models independently?
- Does the agency have sufficient networking capabilities, which are needed to process large quantities of data quickly?[163]

Does the agency have long-term resources available to manage the AI system, such as funding and prioritization in agency goals and strategic planning?

- Is the agency's budget large enough for an AI solution from a trusted vendor or for internal development?[164]
- Can the agency commit long-term resources, including employee time, budget, and prioritization, to maintaining and monitoring the AI solution, as well as to risk mitigation and remediation if necessary? Do the benefits of the AI solution justify these resources?[165]
- Does the agency have capacity and processes in place for monitoring, oversight, and potentially taking the AI out of use?[166]
- Does the agency have flexibility[167] to pivot, pilot, experiment, and iterate beyond what would be used for non-AI tech innovation (due to AI's rapid pace, broad impact, and more autonomous and less transparent nature)?[168]

Is the agency culturally prepared, both internally and externally, for the integration of AI?[169] How do the agency's users, constituents, and employees feel about AI being used in this way?[170] Do they seem prepared for and bought into its integration?

- Is the process or system the AI will be embedded in extremely personal or sensitive (e.g., relating to mental health)? Based on the agency's community engagement research, would using AI in this way feel dehumanizing, disrespectful, or off-putting? In other words, should this decision be left to just humans?[171]
- Is AI targeted only toward constituents in more precarious, high-stakes situations?[172]
- Is AI replacing or complementing a human process? How will humans (either employees or constituents) interact with the system?[173]

METHODOLOGIES

- Consultation with senior leadership that oversee data, privacy, AI, information technology, and/or engineering in the agency
- Consultation with the internal employees who vet, liaise with, and monitor vendors; who develop in-house technology solutions; and who will interact with or use the AI-based system
- Consultation with whoever manages and designs the overall technology development, procurement, and/or review process
- Consultation with any internal stakeholders who intake, document, or manage relevant data sources
- Consultation with senior leadership that oversee long-term budgeting

! RISK ASSESSMENTS[174]

QUESTIONS TO ASK

What is the agency's acceptable level of risk and impact (i.e., the potential consequences it is willing to bear and/or the levels of any metrics that quantify the risk of harm) that AI-powered systems could have on key areas such as civil rights, physical and psychological safety, economic opportunity, educational access, the environment, individuals' or the agency's reputations, and more?[175] How does the agency account for risks that can be quantified as a function of the probability of harm occurring as well as the degree of severity of the harm?[176]

If the agency already has vendor proposals or a specific AI system in mind, has the agency conducted a preliminary risk assessment of the aforementioned risks and others included in frameworks such as the NIST AI RMF?[177]

- What evidence informed the agency's risk assessment?
- What are the data privacy, security, and environmental impacts of the system?[178]
- Are the level and likelihood of these risks within the agency's accepted level of risk tolerance, taking into account any planned mitigations? What are the agency's plans to mitigate the risks it has identified and ensure that they remain within its level of tolerance?

How do the anticipated risks, harms, and other impacts of this AI solution compare to those of the alternatives?

- Specifically, is there concrete evidence that an AI system is likely to be better than alternative solutions in the areas of fairness and bias, both in terms of quantitative metrics and qualitatively in overall impact and outcomes?
- Is the AI system likely to risk significantly more societal harms than other solutions?

What are the agency's tolerance levels for any quantifiable safety, fairness, ethics, or other similar criteria or metrics, and does research suggest that achieving those tolerances is possible given the organization's readiness and technical capabilities?

METHODOLOGIES

- Internal review of risk and impact assessment results with senior leadership, employees handling the development or procurement of the solution, and employees who work directly with constituents in the potential deployment context

QUESTIONS TO ASK LATER IN THE AI DEVELOPMENT LIFECYCLE, BY CRITERION

Agencies are unlikely to be able to answer this next set of questions, which are sorted by the criteria of the *AI Fit Assessment*, *before* deciding on an AI system or use case, which is when the prior list of questions should be initially addressed. However, *after* an agency decides initially to proceed with an AI use case, these follow-up questions should be consulted throughout the rest of the AI development lifecycle alongside the previous set of questions to help the agency decide whether to continue using AI.

EVIDENCE BASE

QUESTIONS TO ASK

Is the now-built AI system (e.g., the specific algorithm, the details of how it is integrated into other agency systems and processes) more effective at meeting the agency's goals than alternative solutions, particularly the status quo and non-AI solutions?

- Was experimentation conducted with other solutions to compare their performance?

Did the AI system's performance meet the agency's original goals? If not, why not? If not, is there a path to remedy?

Did the now-built AI system negatively affect the performance metrics or other guardrails that the agency wanted to maintain while otherwise solving for its problems or goals?

- Did the agency experiment with how the alternatives would affect those metrics? Did the AI solution have a greater adverse impact than the alternatives?

What role does AI play relative to the rest of the system? How does it influence any final outcomes for constituents?[179]

Based on the agency's community engagement research and feedback during the monitoring period, does AI used in this way feel dehumanizing, disrespectful, or off-putting? How do users feel about the solution and how their original problem has been solved?

METHODOLOGIES

- Prior literature, efficacy studies, vendor evaluations, and/or the agency's own preliminary evaluations or experiments
- Review of both vendor-provided information, such as model cards,[180] and third-party evidence
- Community engagement
- User research, interviews, and other forms of community engagement
- Internal discussions and interviews
- Performance metrics, AI evaluations, and external audits

DATA QUALITY

QUESTIONS TO ASK

In design, development, or deployment, have any new gaps or biases in data been revealed?

How will fresh data be integrated into the system during updates, if at all?

If more data has been collected, data has been significantly processed or cleaned since the last time these questions were answered, or data has been more thoroughly explored, reevaluate these questions:

- Has the agency vetted the correctness and precision of the data (if not in its entirety, at least sampled at random)?
- What biases are possible in the data (e.g., selection bias, reporting bias, labeling bias)?[181] Is there an avenue to identify or mitigate those biases?
- Has the agency explored the data to understand its makeup and potential skews or biases in it? Is the data representative of the population at hand and the real world?[182]

METHODOLOGIES

- Exploratory data analysis, exploration, and visualization
- Review of internal documentation and data

ORGANIZATIONAL READINESS

QUESTIONS TO ASK

Reevaluate whether anything about the agency's organizational readiness has changed.

- Have there been budget changes, infrastructure changes, or personnel turnover?
- Are leadership and other stakeholders still bought in?
- Have any constraints around timeline to implementation or the agency's overall goals changed?

METHODOLOGIES

- Internal review

RISK ASSESSMENTS

QUESTIONS TO ASK

Reevaluate the risk assessment and any data privacy, security, and environmental impact assessments.

- Are there mitigation factors designed for both the AI system itself and the larger system it is embedded in (are decisions reversible, humans in the loop, etc.)?
- Is the AI system's performance within the agency's originally desired levels of fairness, safety, etc.?
- Were any harms revealed in experimentation and testing or in deployment? How do they compare to the agency's original risk assessment?
- Do the actual outcomes of the system in the real world seem fair, equitable, safe, and compliant with civil rights requirements when analyzed quantitatively or qualitatively?

METHODOLOGIES

- Internal review of risk assessment results
- Community engagement

Endnotes

1 McKinsey & Company, *The State of AI in Early 2024: Gen AI Adoption Spikes and Starts to Generate Value* (May 30, 2024), https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai [perma.cc/GSS2-JZPU].

2 Bowman Cooper, *More Than Meets the AI: A Look at Some Trends in Federal AI Inventories,* Center for Democracy & Technology (August 15, 2023), https://cdt.org/insights/more-than-meets-the-ai-a-look-at-some-trends-in-federal-ai-inventories/ [perma.cc/T3X7-M7J6].

3 Grant Fergusson, *Outsourced and Automated: How AI Companies Have Taken Over Government Decision-Making,* Electronic Privacy Information Center (2023), https://epic.org/wp-content/uploads/2023/09/FINAL-EPIC-Outsourced-Automated-Report-w-Appendix-Updated-9.26.23.pdf [perma.cc/KY7U-K9BY].

4 Maciej Kuziemski and Gianluca Misuraca, *AI Governance in the Public Sector: Three Tales From the Frontiers of Automated Decision-Making in Democratic Settings,* Telecommunications Policy (July 2020), https://doi.org/10.1016/j.telpol.2020.101976 [perma.cc/Z7CS-BAGG].

5 Google for Developers, *Machine Learning — ML Problem Framing — Understand the Problem* (Accessed September 3, 2024), https://developers.google.com/machine-learning/problem-framing/problem [perma.cc/7TJ9-3HMH].

6 Samuli Laato et al., *AI Governance in the System Development Life Cycle: Insights on Responsible Machine Learning Engineering,* IEEE/ACM 1st International Conference on AI Engineering — Software Engineering for AI (2022), https://dl.acm.org/doi/pdf/10.1145/3522664.3528598 [perma.cc/TY4V-DD5Y].

7 Inioluwa Deborah Raji et al., *The Fallacy of AI Functionality,* ACM Conference on Fairness, Accountability, and Transparency (2022), https://dl.acm.org/doi/pdf/10.1145/3531146.3533158 [perma.cc/9DQ9-X2J5]; AI, Algorithmic, and Automation Incidents and Controversies Repository (Accessed September 3, 2024), https://www.aiaaic.org/aiaaic-repository [perma.cc/X8XE-BNFA].

8 Elizabeth Laird and Hugh Grant-Chapman, *Sharing Student Data Across Public Sectors: Importance of Community Engagement to Support Responsible and Equitable Use,* Center for Democracy & Technology (December 2021), https://cdt.org/wp-content/uploads/2021/12/12-01-2021-Civic-Tech-Community-Engagement-Full-Report-final.pdf [perma.cc/8LUB-GPFA].

9 See notes 3 and 4; Anneke Zuiderwijk et al., *Implications of the Use of Artificial Intelligence in Public Governance: A Systematic Literature Review and a Research Agenda,* Government Information Quarterly (July 2021), https://www.sciencedirect.com/science/article/pii/S0740624X21000137 [perma.cc/69UY-F5Q4]; Yogesh K. Dwivedi et al., *Artificial Intelligence (AI): Multidisciplinary Perspectives on Emerging Challenges, Opportunities, and Agenda for Research, Practice and Policy,* International Journal of Information Management (April 2021), https://www.sciencedirect.com/science/article/abs/pii/S026840121930917X [perma.cc/7JJF-GRGF].

10 See note 4.

11 See note 3.

12 See note 4.

13 White House, *Executive Order on Transforming Federal Customer Experience and Service Delivery to Rebuild Trust in Government* (December 13, 2021), https://www.federalregister.gov/documents/2021/12/16/2021-27380/transforming-federal-customer-experience-and-service-delivery-to-rebuild-trust-in-government [perma.cc/A9CU-PV4Z].

14 See note 2.

15 Executive Department, State of California, *Executive Order N-12-23* (September 6, 2023), https://www.gov.ca.gov/wp-content/uploads/2023/09/AI-EO-No.12-_-GGN-Signed.pdf [perma.cc/35VB-3MD6]; The State of Maryland Executive Department, *Executive Order 01.01.2024.02: Catalyzing the Responsible and Productive Use of Artificial Intelligence in Maryland State Government* (January 8, 2024), https://governor.maryland.gov/Lists/ExecutiveOrders/Attachments/31/EO%2001.01.2024.02%20Catalyzing%20the%20Responsible%20and%20Productive%20Use%20of%20Artificial%20Intelligence%20in%20Maryland%20State%20Government_Accessible.pdf [perma.cc/KG55-UFAX]; Office of the Governor, Commonwealth of Massachusetts, *Executive Order No. 629: Establishing an Artificial Intelligence Strategic Task Force* (February 15, 2024), https://www.mass.gov/doc/executive-order-629/ [perma.cc/D4MJ-CA4U]; State of New Jersey, *Executive Order No. 346* (October 10, 2023), https://nj.gov/infobank/eo/056murphy/pdf/EO-346.pdf [perma.cc/XY7T-B2EZ]; Commonwealth of Pennsylvania Governor's Office, *Executive Order 2023-19 — Expanding and Governing the Use of Generative Artificial Intelligence Technologies Within the Commonwealth of Pennsylvania* (September 20, 2023), https://www.pa.gov/content/dam/copapwp-pagov/en/governor/documents/20230920_EO-2023-19_AI_Final_Executed.pdf [perma.cc/ASF4-5GNS]; State of Rhode Island, *Executive Order 24-06: Artificial Intelligence*

and Data Centers of Excellence (February 29, 2024), https://governor.ri.gov/executive-orders/executive-order-24-06 [perma.cc/5SWP-CE3F]; Commonwealth of Virginia Office of the Governor, *Executive Order Number Thirty (2024): Implementation of Standards for the Safe Use of Artificial Intelligence Across the Commonwealth* (January 18, 2024), https://www.governor.virginia.gov/media/governorvirginiagov/governor-of-virginia/pdf/eo/EO-30.pdf [perma.cc/GT5N-XM75]; State of Washington Office of Governor Jay Inslee, *Executive Order 24-01: Artificial Intelligence* (January 30, 2024), https://governor.wa.gov/sites/default/files/exe_order/24-01%20-%20Artificial%20Intelligence%20%28tmp%29.pdf [perma.cc/87V5-49PX].

16 See note 4.

17 *Ibid.*

18 See note 7.

19 Colin Lecher, *NYC's AI Chatbot Tells Businesses to Break the Law,* The Markup (March 29, 2024), https://themarkup.org/news/2024/03/29/nycs-ai-chatbot-tells-businesses-to-break-the-law [perma.cc/9Q49-GHM2].

20 Official website of the City of New York, *MyCity Chatbot Beta* (2024), https://chat.nyc.gov/ [perma.cc/7FQX-LL44].

21 MIT Management STS Teaching & Learning Technologies, *When AI Gets It Wrong: Addressing AI Hallucinations and Bias* (Accessed September 7, 2024), https://mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias/ [perma.cc/94PN-UHWP].

22 AI, Algorithmic, and Automation Incidents and Controversies Repository, *Thomson Reuters Fraud Detect 'Incorrectly' Identifies Fraud* (January 2024), https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/thomson-reuters-fraud-detect-incorrectly-identifies-fraud [perma.cc/489J-A4D5].

23 *Ibid;* Virginia Eubanks, *A Child Abuse Prediction Model Fails Poor Families: Why Pittsburgh's Predictive Analytics Misdiagnoses Child Maltreatment and Prescribes the Wrong Solutions,* Wired (January 15, 2018), https://www.wired.com/story/excerpt-from-automating-inequality/ [perma.cc/93K5-6C8Z]; AI, Algorithmic, and Automation Incidents and Controversies Repository, *Oregon Drops 'Unfair' Child Abuse Safety at Screening Tool* (December 2022), https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/oregon-dhs-safety-at-screening-tool [perma.cc/Y9RX-5N24].

24 *Ibid* (Eubanks).

25 State of California, *State of California GenAI Guidelines for Public Sector Procurement, Uses and Training* (March 2024), https://www.govops.ca.gov/wp-content/uploads/sites/11/2024/03/3.a-GenAI-Guidelines.pdf [perma.cc/446E-AYKZ].

26 Commonwealth of Virginia, *Policy Standards for the Utilization of Artificial Intelligence by the Commonwealth of Virginia* (June 27, 2024), https://www.vita.virginia.gov/media/vitavirginiagov/it-governance/ea/pdf/Utilization-of-Artificial-Intelligence-by-COV-Policy-Standard.pdf [perma.cc/67RB-2L9U].

27 Maksim Karliuk, *Proportionality Principle for the Ethics of Artificial Intelligence,* AI and Ethics (October 6, 2022), https://link.springer.com/article/10.1007/s43681-022-00220-1 [perma.cc/C78Z-UJZ6]; UNESCO, *Recommendations on the Ethics of Artificial Intelligence* (2022), https://unesdoc.unesco.org/ark:/48223/pf0000381137 [perma.cc/JCG5-3UA8].

28 See note 8.

29 Congressional Research Service, *Cost-Benefit Analysis in Federal Agency Rulemaking* (October 28, 2024), https://crsreports.congress.gov/product/pdf/IF/IF12058 [perma.cc/5UBZ-LGTW].

30 Jan Jöhnk et al., *Ready or Not, AI Comes — An Interview Study of Organizational AI Readiness Factors,* Business & Information Systems Engineering (December 22, 2020), https://link.springer.com/article/10.1007/s12599-020-00676-7 [perma.cc/QRF7-HU8B].

31 National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile* (July 26, 2024), https://www.nist.gov/itl/ai-risk-management-framework [perma.cc/TG9R-ZTM2].

32 Anna Jobin et al., *The Global Landscape of AI Ethics Guidelines,* Nature Machine Intelligence (September 2, 2019), https://www.nature.com/articles/s42256-019-0088-2 [perma.cc/WM9J-NKWK].

33 See note 7; Anna Kawakami et al., *The* Situate AI *Guidebook: Co-Designing a Toolkit to Support Multi-Stakeholder Early-Stage Deliberations Around Public Sector AI Proposals,* Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24) (May 2024), https://dl.acm.org/doi/pdf/10.1145/3613904.3642849 [perma.cc/QU5T-QQK2]; Zhiyuan Wan et al., *How Does Machine Learning Change Software Development Practices?,* IEEE Transactions on Software Engineering (September 1, 2021), https://ieeexplore.ieee.org/document/8812912 [perma.cc/6WWJ-E649].

34 Bob Suh, *When Should You Use AI to Solve Problems?,* Harvard Business Review (February 17, 2021), https://hbr.org/2021/02/when-should-you-use-ai-to-solve-problems [perma.cc/L9TU-28QM].

35 Ian Lundberg et al., *The Origins of Unpredictability in Life Outcome Prediction Tasks,* Proceedings of the National Academy of Sciences (June 4, 2024), https://www.pnas.org/doi/full/10.1073/pnas.2322973121 [perma.cc/CP6E-DMSN].

36 See note 5.

37 See note 7.

38 Miranda Bogen and Amy Winecoff, *Applying Sociotechnical Approaches to AI Governance in Practice,* Center for Democracy & Technology (May 15, 2024), https://cdt.org/insights/applying-sociotechnical-approaches-to-ai-governance-in-practice/ [perma.cc/7UDR-33T4].

39 International Organization for Standardization, *Building a Responsible AI: How to Manage the AI Ethics Debate* (Accessed September 8, 2024), https://www.iso.org/artificial-intelligence/responsible-ai-ethics [perma.cc/D3ZX-K4HE].

40 The following AI lifecycle processes were referenced to develop this one: Organisation for Economic Co-operation and Development AI Principles & Lifecycle [perma.cc/7T6N-69UZ], NIST AI RMF [perma.cc/ESC7-Y6WH], General Services Administration [perma.cc/M6KD-3F24], Center for Data Analytics and Cognition [perma.cc/9WZ5-3M5K], Microsoft's May 2024 Responsible AI Framework [perma.cc/ZNL5-PGQR], case study of machine learning innovation in practice [perma.cc/EN7C-P2GN], Electronic Privacy Information Center's procurement report [perma.cc/SG6N-HD9F], Duke AI Product Management Course [perma.cc/7FJN-DKR9], Google Machine Learning Problem Framing Course [perma.cc/9JLU-KZWG], and City of San Jose AI Review Process [perma.cc/8Q2F-C9G7].

41 Forbes Technology Council, *20 Smart Ways to Create Meaningful Tech Solutions That Meet Real Needs,* Forbes (June 15, 2023), https://www.forbes.com/councils/forbestechcouncil/2023/06/15/20-smart-ways-to-create-meaningful-tech-solutions-that-meet-real-needs/ [perma.cc/TH3Z-9W3G].

42 Bernard Marr, *How to Measure AI Success in Your Organization,* Forbes (June 5, 2024), https://www.forbes.com/sites/bernardmarr/2024/06/05/how-to-measure-ai-success-in-your-organization/ [perma.cc/8HUP-PJN3].

43 IT Modernization Centers of Excellence, *AI Guide for Government: Identifying AI Use Cases in Your Organization,* General Services Administration (Accessed September 12, 2024), https://coe.gsa.gov/coe/ai-guide-for-government/identifying-ai-use-cases-in-your-organization/index.html [perma.cc/WJ3B-9C8U].

44 See note 33 (Wan et al.); IT Modernization Centers of Excellence, *AI Guide for Government: Understanding and Managing the AI Lifecycle,* General Services Administration (Accessed September 12, 2024), https://coe.gsa.gov/coe/ai-guide-for-government/understanding-managing-ai-lifecycle/index.html [perma.cc/WJ3B-9C8U].

45 See notes 8 and 43.

46 See note 43.

47 *Ibid.*

48 Jacqui Ayling and Adriane Chapman, *Putting AI Ethics to Work: Are the Tools Fit for Purpose?,* AI and Ethics (September 12, 2021), https://link.springer.com/article/10.1007/s43681-021-00084-x [perma.cc/RVK8-99KS].

49 See note 38.

50 See note 8; Janis Wong et al., *Key Challenges for the Participatory Governance of AI in Public Administration,* The Alan Turing Institute (2022), https://www.turing.ac.uk/news/publications/key-challenges-participatory-governance-ai-public-administration [perma.cc/5MGY-HFC3].

51 *Ibid.*

52 Peter Cullen and Ilana Golbin, *Evolving to an Effective Algorithmic Impact Assessment,* Information Accountability Foundation (October 2021), https://b1f827.p3cdn1.secureserver.net/wp-content/uploads/2023/08/Evolving-AI-Impact-Assessments-AIA-updated-08072023-1.pdf [perma.cc/MPJ6-PR3S].

53 Anna Kawakami et al., *Studying Up Public Sector AI: How Networks of Power Relations Shape Agency Decisions Around AI Design & Use,* Proceedings of the ACM Conference on Human-Computer Interaction (May 21, 2024), https://arxiv.org/pdf/2405.12458 [perma.cc/8SWS-4DNX]; Hannah Quay-de la Vallee et al., *The Federal Government's Power of the Purse: Enacting Procurement Policies and Practices to Support Responsible AI Use,* Center for Democracy & Technology (April 2024), https://cdt.org/wp-content/uploads/2024/04/2024-04-27-AI-Federal-Procurement-report-1.pdf [perma.cc/764Y-HBBT].

54 This list is not a comprehensive list of public sector use cases of AI; it was developed by reviewing literature on the most promising, effective public sector use cases with minimal or mitigated risk of harm. Some promising use cases may be missing, and use cases in this list may not always be a good fit for an individual situation or agency.

55 See note 25.

56 See note 4.

57 National Association of State Technology Directors, *Artificial Intelligence in State Government IT Operations* (2024), https://higherlogicdownload.s3.amazonaws.com/NASTD/UploadedImages/20b47faa-5f00-40f1-bc5f-7ea8c80514d3/NASTD_AI_Survey_Summary_2024_Final.pdf [perma.cc/DYU8-88XF].

58 See note 4; Michael Chui et al., *Notes From the AI Frontier: Insights From Hundreds of Use Cases,* McKinsey Global Institute (April 2018), https://www.mckinsey.com/west-coast/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Notes%20from%20the%20AI%20frontier%20Applications%20and%20value%20of%20deep%20learning/Notes-from-the-AI-frontier-Insights-from-hundreds-of-use-cases-Discussion-paper.pdf [perma.cc/LG6S-SN2G].

59 See note 2.

60 Eyal Feder-Levy, *Using Generative AI for Community Engagement: Why and How,* International City/County Management Association (June 1, 2024), https://icma.org/articles/pm-magazine/using-generative-ai-community-engagement-why-and-how [perma.cc/L6UH-MM8U].

61 See note 9 (Zuiderwijk et al.).

62 David Freeman Engstrom et al., *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies,* Administrative Conference of the United States (February 1, 2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3551505 [perma.cc/D9N9-GXRZ].

63 See note 60.

64 Thomson Reuters Tax & Accounting, *How Do Different Accounting Firms Use AI?* (June 3, 2024), https://tax.thomsonreuters.com/blog/how-do-different-accounting-firms-use-ai/ [perma.cc/NLV4-XDD3]; Hila Mehr, *Artificial Intelligence for Citizen Services and Government,* Harvard Kennedy School Ash Center for Democratic Governance and Innovation (August 2017), https://creatingfutureus.org/wp-content/uploads/2021/10/Mehr-2017-AIforGovCitizenServices.pdf [perma.cc/HDK9-GYU8].

65 Tom Warren, *Microsoft's New Copilot AI Agents Act Like Virtual Employees to Automate Tasks,* The Verge (May 21, 2024), https://www.theverge.com/2024/5/21/24158030/microsoft-copilot-ai-automation-agents [perma.cc/6F6S-32YP].

66 Begum Karaci Deniz et al., *Unleashing Developer Productivity with Generative AI,* McKinsey Digital (June 27, 2023), https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/unleashing-developer-productivity-with-generative-ai [perma.cc/DEP9-DDLD].

67 See note 58 (Chui et al.).

68 For example, agencies can use semi-supervised anomaly detection methods to flag potential fraud. These algorithms can use both unstructured or unlabeled data and some labeled data of past instances of fraud to detect anomalies in current data. See more examples at https://www.ibm.com/think/topics/machine-learning-for-anomaly-detection [perma.cc/443W-M7K8].

69 International Public Sector Fraud Forum, *The Use of Artificial Intelligence to Combat Public Sector Fraud: Professional Guidance* (February 2020), https://assets.publishing.service.gov.uk/media/5e4545fe40f0b677be5fbd62/Artificial_intelligence_13_Feb.pdf [perma.cc/94K7-JG9J].

70 Darrell M. West, *Using AI and Machine Learning to Reduce Government Fraud,* Brookings Institute (September 10, 2021), https://www.brookings.edu/articles/using-ai-and-machine-learning-to-reduce-government-fraud/ [perma.cc/WZ2E-GAYQ].

71 Natalie Alms, *Using AI to Fight Fraud Is Paying Off, Treasury Says,* Nextgov/FCW (February 28, 2024), https://www.nextgov.com/artificial-intelligence/2024/02/using-ai-fight-fraud-paying-treasury-says/394548/ [perma.cc/6BNH-8X4C].

72 Hannah Quay-de la Vallee, *Combatting Identify Fraud in Government Benefits Programs: Government Agencies Tackling Identity Fraud Should Look to Cybersecurity Methods, Avoid AI-Driven Approaches That Can Penalize Real Applicants,* Center for Democracy & Technology (January 7, 2022), https://cdt.org/insights/combatting-identify-fraud-in-government-benefits-programs-government-agencies-tackling-identity-fraud-should-look-to-cybersecurity-methods-avoid-ai-driven-approaches-that-can-penalize-real-applicant/ [perma.cc/3FSZ-FEYD].

73 Adobe, *Generative AI vs. Other Types of AI* (Accessed September 12, 2024), https://www.adobe.com/products/firefly/discover/generative-ai-vs-other-ai.html [perma.cc/S3AV-NDNV].

74 Edward Graham, *Agencies Are Advancing AI Pilots but Face Scalability Challenges, Report Finds,* Nextgov/FCW (May 9, 2024), https://www.nextgov.com/emerging-tech/2024/05/agencies-are-advancing-ai-pilots-face-scalability-challenges-report-finds/396439/ [perma.cc/B4MY-PN4A].

75 See note 7.

76 Stefano Fantin and Plixavra Vogiatzoglou, *Purpose Limitation by Design as a Counter to Function Creep and System Insecurity in Police Artificial Intelligence,* United Nations Interregional Crime and Justice Research Institute (October 13, 2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3679850 [perma.cc/7TXE-K9FF]; Regulation of the European Parliament and of the Council, *General Data Protection Regulation* (April 27, 2016), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX-:32016R0679&from=EN#d1e1807-1-1 [perma.cc/3R6F-L5MZ].

77 See note 50 (Wong et al.).

78 Paris Will et al., *People Versus Machines: Introducing the HIRE Framework, Artificial Intelligence Review* (May 6, 2022), https://link.springer.com/article/10.1007/s10462-022-10193-6 [perma.cc/HX3H-CLFE].

79 One such evidence scheme is described in the Every Student Succeeds Act. That framework is intended for educational interventions but could be extrapolated to the evidence base for other innovations. Check out the framework on page 19, figure 4: Office of Educational Technology, *Designing for Education With Artificial Intelligence: An Essential Guide for Developers,* U.S. Department of Education (July 2024), perma.cc/NNM8-V9CT.

80 Ben Green, *The Flaws of Policies Requiring Human Oversight of Government Algorithms,* Computer Law & Security Review (July 2022), https://www.sciencedirect.com/science/article/pii/S0267364922000292? [perma.cc/7ADA-VWQ6].

81 Department for Science, Innovation and Technology et al., *Assessing If Artificial Intelligence Is the Right Solution* (June 10, 2019), https://www.gov.uk/guidance/assessing-if-artificial-intelligence-is-the-right-solution [perma.cc/S6KM-YFMB].

82 See note 7.

83 Michael Veale et al., *Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making,* Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (April 21, 2018), https://dl.acm.org/doi/abs/10.1145/3173574.3174014 [perma.cc/5FBF-F4VN].

84 News Staff, *Where Are States Using AI? Survey Says Cybersecurity,* Government Technology (August 16, 2024), https://www.govtech.com/artificial-intelligence/where-are-states-using-ai-survey-says-cybersecurity [perma.cc/32ET-HRLH].

85 Office for Artificial Intelligence, *A Guide to Using Artificial Intelligence in the Public Sector,* UK Government Digital Service (January 2020), https://www.gov.uk/government/publications/a-guide-to-using-artificial-intelligence-in-the-public-sector [perma.cc/Z4MW-RGBW].

86 Carsten Krause, *The Lawler Model for Designing AI Products: A BloombergGPT Case Study,* The CDO Times (September 21, 2023), https://cdotimes.com/2023/09/21/the-lawler-model-for-designing-ai-products-a-bloomberggpt-case-study/ [perma.cc/9C9Z-MGG3].

87 UK Central Digital & Data Office, *Data Ethics Framework* (September 16, 2020), https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020 [perma.cc/TX6A-L56L].

88 *Ibid.*

89 See note 44 (General Services Administration).

90 See note 5.

91 PwC, *Why Most Organizations' Investments in AI Fall Flat,* Harvard Business Review (October 22, 2020), https://hbr.org/sponsored/2020/10/why-most-organizations-investments-in-ai-fall-flat [perma.cc/QUW9-H4AF].

92 Joy Buolamwini and Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification,* Conference on Fairness, Accountability, and Transparency (2018), https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf [perma.cc/CZJ5-3URD].

93 Elizabeth Laird and Hannah Quay-de la Vallee, *Data Ethics in Education & the Social Sector: What Does It Mean and Why Does It Matter?* (February 2021), https://cdt.org/wp-content/uploads/2021/02/2021-02-19-Data-Ethics-and-Ed-and-Social-Sector-FINAL.pdf [perma.cc/Y3ZM-5XQT].

94 Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Punish, and Police the Poor,* Macmillan Publishers (2018).

95 Michael A. Madaio et al., *Learning About Responsible AI On-the-Job: Learning Pathways, Orientations, and Aspirations,* The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24) (June 2024), https://facctconference.org/static/papers24/facct24-103.pdf [perma.cc/T95D-V9RN].

96 See notes 30 and 53 (Kawakami et al.).

97 See note 86.

98 See note 33 (Kawakami et al.).

99 For an explanation of what sociotechnical experts should do in practice, see note 40.

100 See note 30.

101 *Ibid.*

102 See note 4.

103 See note 30.

104 *Ibid.*

105 See note 53 (Kawakami et al.).

106 See notes 31 and 38.

107 See note 31.

108 For more details on risk assessments, guidance on how to conduct them, and additional material that will help agencies answer these guiding questions, see the NIST AI RMF: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [perma.cc/F5XZ-35VG].

109 See note 31.

110 *Ibid.*

111 *Ibid.*

112 Christopher Wilson and Maja van der Velden, *Sustainable AI: An Integrated Model to Guide Public Sector Decision-Making,* Technology in Society (February 2022), https://www.sciencedirect.com/science/article/pii/S0160791X22000677/ [perma.cc/Q4PW-ABAT].

113 See note 31.

114 *Ibid.*

115 See note 8.

116 Memorandum for the Heads of Executive Departments and Agencies, *Transparency and Open Government,* White House (January 21, 2009), https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government [perma.cc/W9RU-C8TX].

117 Dan Bateyko et al., *CDT Comments on OMB Draft Guidance for Agency Use of AI,* Center for Democracy & Technology (December 5, 2023), https://cdt.org/insights/cdt-comments-on-omb-draft-guidance-for-agency-use-of-ai/ [perma.cc/3TVF-GTSA].

118 See note 8.

119 *Ibid.*

120 *Ibid.*

121 *Ibid.*

122 Hugh Grant-Chapman and Hannah Quay-de la Vallee, *Making Government Data Publicly Available: Guidance for Agencies on Releasing Data Responsibly* (August 2022), https://cdt.org/wp-content/uploads/2022/08/2022-08-29-CDT-Data-Guidance-accessible.pdf [perma.cc/69FB-R926].

123 See note 78.

124 See note 80.

125 See note 81.

126 See note 7.

127 Vincent J. Straub et al., *Artificial Intelligence in Government: Concepts, Standards, and a Unified Framework,* Government Information Quarterly (October 2023), https://www.turing.ac.uk/sites/default/files/2023-11/straub-et-al-2022-ai-in-gov.pdf [perma.cc/GHS8-Q2FL]; see notes 9 (Dwivedi et al.) and 83.

128 See notes 4 and 7.

129 See note 30.

130 See note 58 (Chui et al.).

131 See note 7 (Raji et al.).

132 Arif Kornweitz, *A New AI Lexicon: Function Creep,* AI Now Institute (August 4, 2021), https://ainowinstitute.org/publication/a-new-ai-lexicon-function-creep [perma.cc/Z47T-9Z7L].

133 See note 4.

134 See note 33 (Kawakami et al.).

135 Government of Canada, *Algorithmic Impact Assessment Tool* (Accessed September 18, 2024), https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html [perma.cc/GA5C-L69D]; see note 9 (Zuiderwijk et al.).

136 *Ibid.*

137 See note 83.

138 See note 4.

139 Nari Johnson et al., *The Fall of an Algorithm: Characterizing the Dynamics Toward Abandonment,* FAccT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (June 2024), https://dl.acm.org/doi/10.1145/3630106.3658910 [perma.cc/WD6F-FGRB].

140 Margaret Mitchell et al., *Model Cards for Model Computing,* FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency (2019), https://dl.acm.org/doi/10.1145/3287560.3287596 [perma.cc/CN5D-FSWF].

141 See note 8.

142 See note 87.

143 See note 6.

144 Organisation for Economic Co-operation and Development, *Scoping the OECD AI Principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO),* OECD Digital Economy Papers (2019), https://doi.org/10.1787/20716826 [perma.cc/SV85-LJFM].

145 Joy Buolamwini, *Unmasking the Bias in Facial Recognition Algorithms,* MIT Sloan School of Management (December 13, 2023), https://mitsloan.mit.edu/ideas-made-to-matter/unmasking-bias-facial-recognition-algorithms [perma.cc/6G52-CS3Q].

146 See note 44 (General Services Administration).

147 *Ibid.*

148 See notes 5, 9 (Dwivedi et al.), and 33 (Wan et al.).

149 Michael Yang, *Dispelling Myths About Artificial Intelligence for Government Service Delivery,* Center for Democracy & Technology (January 25, 2023), https://cdt.org/insights/dispelling-myths-about-artificial-intelligence-for-government-service-delivery/ [perma.cc/4DV8-BW7L].

150 See note 91.

151 See note 31.

152 See note 91.

153 See note 5.

154 *Ibid.*

155 See note 33 (Kawakami et al.).

156 See note 53 (Kawakami et al.).

157 See note 38 for a detailed explanation of what sociotechnical experts should do in practice.

158 See note 95.

159 See note 34.

160 *Ibid.*

161 See note 91.

162 See note 30.

163 *Ibid.*

164 Patrick Mikalef et al., *Enabling AI Capabilities in Government Agencies: A Study of Determinants for European Municipalities,* Government Information Quarterly (October 2022), https://www.sciencedirect.com/science/article/pii/S0740624X21000320 [perma.cc/MAZ6-GA7N].

165 See note 5.

166 See note 91.

167 Friso Selten and Bram Klievink, *Organizing Public Sector AI Adoption: Navigating Between Separation and Integration,* Government Information Quarterly (March 2024), https://www.sciencedirect.com/science/article/pii/S0740624X23000850 [perma.cc/SLX6-7PWQ].

168 *Ibid.*

169 See note 30.

170 See note 4.

171 *Ibid.*

172 *Ibid.*

173 See note 52.

174 See note 31.

175 *Ibid.*

176 *Ibid.*

177 *Ibid.*

178 See note 112.

179 See note 4.

180 See note 140.

181 See note 144.

182 See note 145.

cdt.org

cdt.org/contact

Center for Democracy & Technology
1401 K Street NW, Suite 200
Washington, D.C. 20005

202-637-9800

@CenDemTech

CENTER FOR
DEMOCRACY
& TECHNOLOGY