ADMINISTRATIVE CONFERENCE OF THE UNITED STATES

Agency Use of Artificial Intelligence

Ad Hoc Committee on Agency Use of Artificial Intelligence

Draft Statement for Ad Hoc Committee | November 30, 2020

Artificial intelligence (AI) techniques are changing how government agencies do their work.[1] Advances in AI hold out the promise of lowering the cost of completing government tasks and improving the quality, consistency, and predictability of agency decisions. But enhanced agency use of AI also raises concerns about the discretion being vested in AI systems and the extent to which those systems are exercising authority that ought to be handled by human officials.

Consistent with its statutory mission to promote efficiency, participation, and fairness in administrative processes,[2] the Administrative Conference offers this Statement to identify issues of which agencies should be mindful when adopting or modifying AI systems. The Statement

Commented [A1]: I'm concerned that defining AI narrowly as "software or hardware" obscures the importance of other vital components of a system – some of which are discussed below, such as data and human-centered processes. It also has the tendency to confuse the applicability of agency responsibilities related to those components. Further, this definition focuses on "technology," whereas there is a roughly equal tendency throughout the document to refer to "AI techniques," which suggests a disconnect. I think the better and more consistent conceptualization of AI is as an "information system," as defined at 44 U.S.C. 3502 and in A-130 and which serves as the basis of many agency information and IT responsibilities in statute and policy. I understood from the previous meeting that, because of the difficulty identifying a single, concrete definition, we were going to focus on giving illustrative examples.

[1] ~~The National Institute of Standards and Technology has offered the following basic definition of AI:~~

~~AI technologies and systems are considered to comprise software [or] hardware that can learn to solve complex problems, make predictions or undertake tasks that require human-like sensing (such as vision, speech, and touch), perception, cognition, planning, learning, communications, or physical action. Examples are wide-ranging and expanding rapidly.~~ There is no universally accepted definition of "Artificial Intelligence," and the rapid state of evolution in the field, as well as the proliferation of use cases, makes coalescing around any such definition difficult. Generally speaking, AI systems tend to have characteristics such as the ability to learn to solve complex problems, make predictions, or undertake tasks that heretofore have relied on human decisionmaking or intervention. There are many illustrative examples of AI that can help frame the issue for the purpose of this statement. They include, but are not limited to, AI assistants, computer vision systems, biomedical research, unmanned vehicle systems, advanced game-playing software, and facial recognition systems as well as application of AI in both Information Technology (IT) and Operational Technology (OT).

Formatted: Indent: Left: 0", Right: 0"

~~NAT'L INST. OF STANDARDS & TECH., U.S. LEADERSHIP IN AI: A PLAN FOR FEDERAL ENGAGEMENT IN DEVELOPING TECHNICAL STANDARDS AND RELATED TOOLS 7–8 (Aug. 9, 2019). The Administrative Conference adopts that definition for purposes of this statement.~~

[2] *See* 5 U.S.C. § 591.

draws on a pair of reports commissioned by the Conference,[3] as well as the input of AI experts from government, academia, and the private sector.

The issues highlighted in this Statement are relevant to an array of agency personnel. To minimize the risk of unforeseen problems involving an AI system, the agency should, throughout the system's lifespan, solicit input about the system from an array of offices—including, at a minimum, the legal, policy, financial, human resources, and technology offices.

1. Transparency

Agencies' efforts to ensure transparency in connection with their AI systems can serve many valuable goals. When agencies set up processes to ensure transparency in their AI systems, they should publicly identify the processes' goals and the rationales behind them. For example, an agency might prioritize transparency in the service of legitimizing its AI systems, facilitating internal or external review of its AI-based decisionmaking, or coordinating its activities. Different AI systems are likely to satisfy some transparency goals more than others. Where possible, agencies should use metrics to measure the performance of their AI-transparency processes.

In setting transparency goals, agencies should consider to whom they should be transparent. For instance, depending on the nature of its operations, an agency might prioritize transparency to the public, courts, Congress, or its own officials.

The appropriate level or nature of transparency and interpretability in an agency's AI systems will also depend on context. In some contexts, such as adjudication, reason-giving requirements may call for a high degree of transparency and interpretability from the agency regarding how an AI system functions. In other contexts, such as enforcement, an agency's legitimate interests in preventing gaming or adversarial learning by regulated parties could

[3] DAVID FREEMAN ENGSTROM, DANIEL E. HO, CATHERINE M. SHARKEY, & MARIANO-FLORENTINO CUÉLLAR, GOVERNMENT BY ALGORITHM: ARTIFICIAL INTELLIGENCE IN FEDERAL ADMINISTRATIVE AGENCIES (2020), https://www.acus.gov/sites/default/files/documents/Government%20by%20Algorithm.pdf; Cary Coglianese, *A Framework for Governmental Use of Machine Learning* (Oct. 2020), https://www.acus.gov/sites/default/files/documents/Coglianese%20Report%20-%20A%20Framework%20for%20Governmental%20Use%20of%20Machine%20Learning.pdf (draft report for Administrative Conference of the United States).

militate against providing too much information (or specific types of information) to the public about the AI system's processes. In each context, agencies should consider whether particular laws or policies governing disclosure of information apply.

In selecting and using AI techniques, agencies should be cognizant of the degree to which a particular AI system can be made transparent to appropriate people and entities, including the general public. There may exist tradeoffs between explainability and accuracy in AI systems, so that transparency and interpretability might sometimes weigh in favor of choosing simpler AI models. The appropriate balance between explainability and accuracy will depend on the agency's circumstances and priorities.

The proprietary nature of some AI systems may also affect the extent to which they can be made transparent. When an agency's AI system relies on proprietary technologies or algorithms the agency does not own, the agency and the public may have only limited access to the information needed to understand the AI technique. Agencies should strive to anticipate such circumstances and address them appropriately, such as by working with outside providers to ensure they will be able to share sufficient information about such a system.

2. Harmful Bias

At their best, AI systems can help agencies identify and reduce the impact of unwanted human biases.[4] Yet they can also unintentionally create or exacerbate those biases by encoding and deploying them at scale. In deciding whether and how to deploy an AI system, therefore, agencies should carefully evaluate the biases that might result from the use of the AI system as well as the biases that might result from alternative systems that rely on human actors (such as an incumbent system that the AI system would augment or replace). Because different types of bias

[4] The term *bias* has a technical meaning in the machine learning literature related to model characteristics. Under some circumstances, increasing bias (roughly the error of the average prediction) can improve system performance, if it reduces the risk of overfitting. Here, the Administrative Conference uses the term more generally to refer to common or systematic errors in decision making, especially those implicating normative concerns related to fairness and equal treatment.

pose different types of harms, the outcome of the evaluation will depend on the agency's unique circumstances and priorities and the consequences posed by those harms in that context.

AI systems can be biased because of their reliance on data reflecting historical human biases or because of their designs. Biases in AI systems can increase over time through feedback, which can occur if the use of a biased AI system leads to systematic errors in categorizations, which are then reflected in the data set or data environment the system uses to make future predictions. Agencies should be mindful of the interdependence of the models, metrics, and data that underpin AI systems.

Identifying biases in AI systems can pose challenges, as when the bias affects a particular population but information about which individuals are in that population is not directly available. To identify and mitigate such biases, agencies should, to the extent practical, consider whether other data or methods are available.

Data science techniques for identifying and mitigating biases in AI systems are developing. Agencies should stay up to date on developments in the field of AI, particularly on algorithmic fairness; establish processes to ensure that people with diverse perspectives are able to inspect AI systems and their decisions for indications of harmful bias; test AI systems in environments resembling the ones in which they will be used; and make use of internal and external processes for evaluating the risks of bias in AI systems.

3. Technical Capacity

AI systems can help agencies conserve resources, but they can also require substantial investments of human and financial capital. Agencies should carefully evaluate the short- and long-term costs and benefits of an AI system before committing significant resources to it. Each agency should also ensure it has access to the technical expertise required to make informed decisions about the type of AI systems it requires, how to integrate those systems into its operations, and how to oversee, maintain, and update those systems.

Given the data science field's ongoing and rapid development, agencies should consider cultivating an AI-ready workforce, including through recruitment and training efforts that

emphasize AI skills. When agency personnel lack the skills to develop, procure, or maintain an AI system that meets the agency's needs, the agency should consider other means of expanding its technical expertise, including by relying on tools such as the Intergovernmental Personnel Act,[5] prize competitions, or cooperative research and development agreements with private institutions or universities.

4. Obtaining AI Systems

Decisions about whether or how to obtain an AI system can involve important trade-offs. Buying an AI system from an external source might allow the agency to acquire a more sophisticated tool than it could design on its own, access that tool sooner, and save some of the up-front costs associated with developing the technical capacity needed to design an AI system. Creating an AI tool within the agency, by contrast, might yield a tool that is better tailored to the agency's particular tasks and policy goals. Creating an AI system within the agency can also facilitate development of internal technical capability, which can yield benefits over the lifetime of the AI system and in other technological tasks the agency may confront.

Certain government offices are available to help agencies with decisions and actions related to technology.[6] Agencies should make appropriate use of these resources when obtaining an AI system.

5. Data

AI systems require data, often in vast quantities. An agency should consider whether it has, or can obtain, data that appropriately reflects conditions similar to the ones the agency's AI systems will address in practice; whether the agency has the resources to render the data into a

[5] 5 U.S.C. §§ 3371–76.

[6] Within the General Services Administration, for example, the office called 18F routinely partners with government agencies to help them build and buy technologies. Similarly, the United States Digital Service has a staff of technologists whose job is to help agencies build better technological tools. While the two entities have different approaches—18F acts more like an information intermediary and the Digital Service serves as an alternative source for information technology contracts—both could aid agencies with obtaining, developing, and using different AI techniques.

format that can be used by the agency's AI systems; and how the agency will maintain the data and link it to the agency's AI systems without compromising security or privacy.

6. Privacy

Agencies have a responsibility to protect privacy with respect to personally identifiable information in AI systems no less than in other aspects of agency operation. In a narrow sense, this responsibility demands that agencies comply with requirements related to, for instance, transparency, due process, accountability, and information quality and integrity established by the Privacy Act of 1974, Section 208 of the E-Government Act of 2002, and other laws and policies.[7] More broadly, agencies should recognize and appropriately manage privacy risks posed by an AI system. Agencies should consider privacy risks throughout the entire development life cycle of an AI system and assess those risks, as well as associated controls, on an ongoing basis. The Office of Management and Budget and the National Institute of Standards and Technology have developed a R~~r~~isk M~~m~~anagement F~~f~~rameworks ~~that agencies may find useful~~for agencies to utilize in implementing AI systems.[8]

7. Security

Agencies should consider the possibility that AI systems might be manipulated, fooled, evaded, and misled, including through manipulation of training data and exploitation of model sensitivities. An agency must ensure not only that its data is secure, but also that its AI systems are trained on that data in a secure manner, make forecasts based on that data in a secure manner, and otherwise operate in a secure manner. Agencies should continuously consider and evaluate the safety and security of AI systems, including resilience to vulnerabilities, manipulation, and other malicious exploitation.

[7] *See, e.g.* 5 U.S.C. § 552a(e), (g), & (p); 44 U.S.C. § 3501 note.

[8] *See~~, e.g.,~~* ~~Nat'l Inst. of Standards & Tech., *NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management, Version 1.0* (Jan. 16, 2020),~~ Nat'l Inst. of Standards & Tech. Special Publication SP-800-37 revision 2, *Risk Management Framework for Information Systems and Organizations: A System Lifecycle Approach for Security and Privacy* (Dec. 2018); Office of Mgmt. & Budget, Circular A-130, *Managing Information as a Strategic Resource* (July 28, 2016). *See also* Nat'l Inst. of Standards & Tech., *NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management, Version 1.0* (Jan. 16, 2020).

Commented [A2]: In the previous draft, these subject descriptions were meant to align (though imperfectly) with other sections in the statement, which are less obvious in this draft. These descriptions were not meant to represent the universe of requirements. It would be a mischaracterization of agency responsibilities to suggest that these are the only requirements related to privacy.

Commented [A3]: "May find useful" mischaracterizes the responsibility to use the Risk Management Framework (RMF), which is a proper noun. A-130 makes agency use of the RMF, as fleshed out in SP-800-37, mandatory. The NIST Privacy Framework, however, is not mandatory for agencies.

Formatted: Font: 9 pt

8. Decisional Authority

Agencies should be mindful that most AI systems will involve human beings in a range of capacities—as operators, customers, overseers, policymakers, or interested members of the public. Accordingly, any decision to deploy an AI system should account for the human tendencies and preferences of humans in those roles.

Human factors may sometimes undercut the value of using AI systems to make certain determinations. There is a risk, for example, that human operators will devolve too much responsibility to AI systems and fail to detect cases where the AI systems yield inaccurate or unreliable determinations. That risk may be tolerable in some settings—such as when the AI system has recently been shown to perform significantly better than alternatives—but intolerable in others.

Similarly, if agency personnel come to rely reflexively on algorithmic results in exercising discretionary powers, use of an AI system could have the practical effect of curbing the exercise of agency discretion or shifting it from the person who is supposed to be exercising it to the system's designer. Agencies should beware of such potential shifts of practical authority and take steps to ensure that appropriate officials have the knowledge and power to be accountable for decisions made or aided by AI techniques.

Finally, there may be some circumstances where, for reasons wholly apart from decisional accuracy, an agency may wish to have a decision be made by a human being, even if the law does not require it. In some contexts, accuracy and fairness are not the only relevant values at stake, and an AI system may be difficult to sustain if human beings perceive it as unfair, inhumane, or otherwise unsatisfactory.[9]

9. Oversight

[9] *Cf.* Admin. Conf. of the U.S., Recommendation 2018-3, *Electronic Case Management in Federal Administrative Adjudication*, 83 Fed. Reg. 30,686 (June 29, 2018) (suggesting, in the context of case management systems, that agencies consider implementing electronic systems only when they conclude that doing so would lead to benefits without impairing either the objective "fairness" of the proceedings or the subjective "satisfaction" of those participating in those proceedings).

It is essential that agencies' AI systems be subject to appropriate and regular oversight throughout their lifespans. There are two general categories of oversight: external and internal. An agency's mechanisms of internal oversight will be shaped by the demands of external oversight. And the more effective an agency's internal oversight mechanisms, the better it is likely to fare with external oversight. An agency should be cognizant of both forms of oversight in making decisions about its AI systems.

External oversight of agency use of AI systems can come from a variety of government sources, including inspectors general, the Government Accountability Office, and Congress. Courts can also play an important role in external oversight of agency uses of AI systems. Agency uses of AI systems might lead to litigation in a number of circumstances. Those affected by an agency's use of an AI system might, for example, allege that use of the system violates their right to procedural due process.[10] Or they might allege that the AI system's determination violated the Administrative Procedure Act (APA) because it was arbitrary and capricious.[11] When an AI system narrows the discretion of agency personnel, or fixes or alters the legal rights and obligations of people subject to the agency's action, affected people or entities might also sue on the ground that the AI system is a legislative rule adopted in violation of the APA's requirement that legislative rules go through the notice-and-comment process.[12] Agencies should consider these different forms of potential external oversight as they are making and documenting decisions about AI systems and as they are developing processes for making those decisions.

Agencies should also develop their own, internal evaluation and oversight mechanisms for their uses of AI systems. Successful internal oversight requires advance and ongoing planning and consultation with the various offices in an agency that will be affected by the agency's use of an AI system, including its legal, policy, financial, human resources, and

[10] Courts would analyze such challenges under the three-part balancing framework from *Mathews v. Eldridge*, 424 U.S. 319, 335 (1976).

[11] *See* 5 U.S.C. § 706(2)(A). Courts would review such challenges under the standard set forth in *Motor Vehicle Manufacturers Ass'n v. State Farm Mutual Automobile Insurance Co.*, 463 U.S. 29, 43 (1983).

[12] *See* 5 U.S.C. § 553(b)–(c).

technology offices. An agency's oversight plan should address how the agency will pay for its oversight mechanisms and how it will respond to what it learns from its oversight.

Agencies should establish a protocol for regularly evaluating AI systems throughout the systems' lifespans. That is particularly true if a system or the circumstances in which it is deployed are liable to change over time, since, in that case, review and explanation of the system's functioning at one stage of development or use may become outdated due to changes in the system's underlying models. To enable that type of oversight, agencies should monitor and keep track of the data being used by their AI systems, as well as how the systems use that data. Agencies may also wish to secure input from members of the public or private evaluators to improve the likelihood that they will identify defects in their AI systems.

To make their oversight systems more effective, agencies should clearly define goals for their AI systems. The relevant question for oversight purposes will often be whether the AI system outperforms alternatives, which may require the agency to benchmark the system against the status quo or some hypothetical state of affairs.

Finally, AI systems can affect how agency staff do their jobs, particularly as agency personnel grow to trust and rely on the systems. In addition to evaluating and overseeing their AI systems, agencies should pay close attention to how agency personnel interact with those systems.