ADMINISTRATIVE CONFERENCE OF THE UNITED STATES

Agency Use of Artificial Intelligence

Ad Hoc Committee on Agency Use of Artificial Intelligence

Proposed Statement | December 16, 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 agencies' decisions. But agencies' uses of AI also raise concerns about the discretion being vested in AI systems and the extent to which those systems are exercising authority previously exercised 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 agencies should consider when adopting or modifying AI systems and developing practices and procedures for their use and regular monitoring. The Statement draws on a pair of reports commissioned by the Conference,[3] as well as the input of AI experts from government,

[1] 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. *See, e.g.*, John S. McCain National Defense Authorization Act for Fiscal Year 2019, Pub. L. No. 115-232, § 238(g), 132 Sta. 1636, 1697–98 (2018) (using one definition of AI); 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) (offering a different definition of AI). 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 decision making 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 and operational technology.

[2] *See* 5 U.S.C. § 591.

[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).

academia, and the private sector (some ACUS members) provided at meetings of the ad hoc committee of the Administrative Conference that proposed this Statement.

The issues addressed in this Statement implicate matters involving law, policy, finances, human resources, and technology. To minimize the risk of unforeseen problems involving an AI system, agencies should, throughout an AI system's lifespan, solicit input about the system from the offices that oversee these matters. Agencies should also keep in mind the need for public trust in their practices and procedures for use and regular monitoring of AI technologies.

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 consider publicly identifying the processes' goals and the rationales behind them. For example, agencies might prioritize transparency in the service of legitimizing its AI systems, facilitating internal or external review of its AI-based decision making, or coordinating its AI-based 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, agencies might prioritize transparency to the public, courts, Congress, or their own officials.

The appropriate level or nature of transparency and interpretability in agencies' AI systems will also depend on context. In some contexts, such as adjudication, reason-giving requirements may call for a higher degree of transparency and interpretability from agencies regarding how their AI systems function. In other contexts, such as enforcement, agencies' legitimate interests in preventing gaming or adversarial learning by regulated parties could militate against providing too much information (or specific types of information) to the public about AI systems' processes. In every 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 specific context, including agencies' circumstances and priorities.

The proprietary nature of some AI systems may also affect the extent to which they can be made transparent. When agencies' AI systems rely on proprietary technologies or algorithms the agencies do not own, the agencies and the public may have limited access to the information about the AI techniques. 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. Agencies should not enter into contracts to use proprietary AI systems unless they are confident that actors both internal and external to the agencies will have adequate access to information about the systems.

2. Harmful Bias

At their best, AI systems can help agencies identify and reduce the impact of unwanted 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, agencies should carefully evaluate the harmful biases that might result from the use of the AI system as well as the biases that might result from alternative systems (such as an incumbent system that the AI system would augment or replace). Because different types of bias pose different types of harms, the outcome of the evaluation will depend on agencies' unique circumstances and priorities and the consequences posed by those harms in those contexts.

[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 concerns related to fairness and equal treatment.

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. That can occur, for example, 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 harmful biases in AI systems can pose challenges, as when the bias affects a particular population but information about those 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. Agencies should periodically examine and refresh AI algorithms and other protocols to ensure that they remain sufficiently current and reflect new information and circumstances relevant to the functions they perform.

Data science techniques for identifying and mitigating harmful 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 personnel that reflect various disciplines and relevant 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 harmful bias in AI systems and for identifying such bias.

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. Agencies should also ensure they have access to the technical expertise required to make informed decisions about the type of AI systems they require; how to integrate those systems into their 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 AI systems that meets agencies' needs, agencies should consider other means of expanding their technical expertise, including by relying on tools such as the Intergovernmental Personnel Act,[5] prize competitions, cooperative research and development agreements with private institutions or universities, and consultation with external technical advisors and subject-matter experts.

4. Obtaining AI Systems

Decisions about whether to obtain an AI system can involve important trade-offs. Obtaining AI systems from external sources might allow agencies to acquire more sophisticated tools than they could design on their own, access those tools sooner, and save some of the up-front costs associated with developing the technical capacity needed to design AI systems.[6] Creating AI tools within agencies, by contrast, might yield tools that are better tailored to the agencies' particular tasks and policy goals. Creating AI systems within agencies can also facilitate development of internal technical capability, which can yield benefits over the lifetime of the AI systems and in other technological tasks the agencies may confront.

Certain government offices are available to help agencies with decisions and actions related to technology.[7] Agencies should make appropriate use of these resources when obtaining an AI system. Agencies should also consider the cost and availability of the technical support

[5] 5 U.S.C. §§ 3371–76.

[6] Agencies may also obtain AI systems that are embedded in commercial products. The considerations applicable to such embedded AI systems should reflect the fact that agencies may have less control over their design and development.

[7] 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.

necessary to ensure that an AI system can be maintained and updated in a manner consistent with its expected life cycle and service mission.

5. Data

AI systems require data, often in vast quantities. Agencies should consider whether they have, or can obtain, data that appropriately reflects conditions similar to the ones the agencies' AI systems will address in practice; whether the agencies have the resources to render the data into a format that can be used by the agencies' AI systems; and how the agencies will maintain the data and link it to their AI systems without compromising security or privacy. Agencies should also review and consider statutes and regulations that impact their uses of AI as a potential consumer of data.

6. Privacy

Agencies have a responsibility to protect privacy with respect to personally identifiable information in AI systems. 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 applicable laws and policies.[8] More broadly, agencies should recognize and appropriately manage privacy risks posed by an AI system. Agencies should consider privacy risks throughout the entire life cycle of an AI system from development to retirement and assess those risks, as well as associated controls, on an ongoing basis. In designing and deploying AI systems, agencies should consider using relevant privacy risk management frameworks developed through open, multi-stakeholder processes.[9]

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

[9] *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).

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. Agencies must ensure not only that their data is secure, but also that their 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. In designing and deploying AI systems, agencies should consider using relevant voluntary consensus standards and frameworks developed through open, multi-stakeholder processes.[10]

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. 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 agencies' 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

[10] *See, e.g.*, NAT'L INST. FOR STANDARDS & TECH., FRAMEWORK FOR IMPROVING CRITICAL INFRASTRUCTURE CYBERSECURITY (Apr. 16, 2018).

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, agencies may wish to have decisions be made by human beings, even if the law does not require it. In some contexts, accuracy and fairness are not the only relevant values at stake, and AI systems may be difficult to sustain if human beings perceive them as unfair, inhumane, or otherwise unsatisfactory.[11]

9. Oversight

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. Agencies' mechanisms of internal oversight will be shaped by the demands of external oversight. Agencies should be cognizant of both forms of oversight in making decisions about their AI systems.

External oversight of agencies' uses of AI systems can come from a variety of government sources, including inspectors general, externally-facing ombuds, the Government Accountability Office, and Congress. In addition, because agencies' uses of AI systems might lead to litigation in a number of circumstances, courts can also play an important role in external oversight. 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.[12] Or they might allege that the AI system's determination violated the Administrative Procedure Act (APA) because it was

[11] *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).

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

arbitrary and capricious.[13] 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.[14] Agencies should consider these different forms of potential external oversight as they are making and documenting decisions and the underlying processes for these AI systems.

Agencies should also develop their own internal evaluation and oversight mechanisms for their AI systems, both for initial approval of an AI system and for regular oversight of the system. 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, internally-facing ombuds, and technology offices. Agencies' oversight plans should address how the agencies will pay for their oversight mechanisms and how they will respond to what they learn from their 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. In these instances, 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

[13] *See* 5 U.S.C. § 706(2)(A). Courts would likely 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).

[14] *See* 5 U.S.C. § 553(b)–(c).

system outperforms alternatives, which may require agencies to benchmark their systems against the status quo or some hypothetical state of affairs.

Finally, AI systems can affect how agencies' staffs 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.