



Towards a Framework of Institutional Trust for AI Regulatory Enforcement

Michael Karanicolas, Executive Director, UCLA Institute for Technology, Law, and Policy (karanicolas@law.ucla.edu)

In recent years, an increasing number of government agencies have incorporated algorithmic systems into their regulatory enforcement processes. The efficacy gains from these systems are easy to imagine, especially with growing public demands from the administrative state to do more with less. They may also present opportunities to expand the processing power of agencies and boost research and analytical capabilities in unprecedented ways. However, the incorporation of AI-powered tools into core regulatory enforcement functions also raises a number of difficult questions, especially regarding the fairness of these systems and their broader impacts on decision-making and enforcement processes. This brief explores these concerns, and raises considerations that stakeholders should bear in mind when evaluating and addressing them.

Agencies Must Consider Ramifications of AI Use

The use of AI tools for official decision-making raises several core questions surrounding transparency, accountability, and procedural fairness. At a time when trust in public institutions has significantly declined, there is great potential for AI systems to further erode the relationship between the people and the administrative state.

Troublingly, agencies' enthusiasm for harnessing AI runs the risk of ignoring these concerns regarding AI ethics and transparency. Efforts to track how and where AI is implemented across administrative agencies are patchy at best, relying on interviews and anecdotal evidence rather than comprehensive mapping or assessment. This incomplete information presents a significant obstacle to cultivating effective oversight around these tools and establishing best practices for appropriate safeguards or policies regarding the use of AI for regulatory enforcement. Poor information generally results from insufficient coordination across governmental bodies experimenting with AI tools.

Evaluating the Role of AI in a Decision-Making Context

Many discussions regarding the pros and cons of AI in administrative decision-making focus on the weaknesses of human decision-makers, such as their inconsistency and tendencies towards bias. However, these comparisons ultimately miss the point. Our system of due process evolved in response to the specific challenges posed by our current, human-centric model of decision-making. Adding AI to the mix presents an entirely new set of challenges which our current standards of due process may be ill-equipped to address.

The use of AI in administrative decision-making raises fundamental questions regarding whether these technologies can achieve the same underlying values that our system of administrative law aims at. Personal dignity is one example, where it is unclear whether people will accept having major decisions about their lives made by machines. Fairness in the decision-making process is only part of the equation, and perceptions of fairness are an equally important value to upholding the legitimacy of the system. **The essential question is not whether we can effectively port existing standards over to an AI context, but rather what safeguards – if any – are capable of delivering the same underlying values.**

Addressing Governmental Uses of AI

Responsible and accountable government mandates an open and public conversation about the use of AI to perform critical tasks, particularly those related to regulatory enforcement, where due process concerns are paramount. Ultimately, the goal of administrative agencies across all levels of government should be to craft an effective and harmonized decision-making framework regarding the use of AI across the public sector.

Effective tracking and coordination is an essential first step to providing a complete picture of the existing and potential uses of these technologies. A necessary second step is **developing a robust process for evaluating the successes and failures of AI systems** in order to inform their future implementation.



Figure 1: AI RMF (NIST 2024)

Crafting Regulatory Standards for Governmental Uses of AI

To date, the National Institute of Standards and Technology (NIST)'s [Artificial Intelligence Risk Management Framework \(AI RMF\)](#), illustrated in **Figure 1**, is the most comprehensive attempt at crafting regulatory standards for governmental use of AI tools. **The AI RMF establishes a taxonomy of potential risks stemming from the use of AI**, including defining challenges related to their reliability, accuracy, robustness, security, accountability, privacy, fairness, and bias. This framework also provides a model risk assessment structure, which tracks and measures potential risks, and responds to them as they emerge.

While the AI RMF provides a valuable starting point, it still leaves significant conceptual heavy-lifting in the hands of individual agencies. AI-related risks are ultimately contextual determinations, which depend heavily on specific use cases. This means that **the AI RMF relies on a sense of collective responsibility for managing the impacts of AI**, as well as a willingness among relevant agencies to ask resource-intensive and difficult questions about tradeoffs resulting from these various use cases.