Mathematica
Progress Together

Response to the Request for Information on

National Priorities for Artificial Intelligence

July 7, 2023

Submitted to:

Office of Science and Technology Policy
Executive Office of the President
1650 Pennsylvania Ave., NW
Washington, DC 20504

Attention: Stacy Murphy
Operations Manager

Submitted by:

Mathematica
1100 First Street, NE, 12th Floor
Washington, DC 20002-4221
Phone: (202) 484-9220
Fax: (609) 228-4958

Introduction

Mathematica is a nonpartisan research and data analytics organization with a mission to improve public well-being. The rapid development and deployment of artificial intelligence (AI) creates both enormous opportunities and risks to public well-being. Mathematica's deep, interdisciplinary expertise—including our experience using AI tools—positions us to anticipate many opportunities and risks alike, and to consider the ways policy might support those opportunities and mitigate those risks.

We applaud the interest of the Office of Science and Technology Policy (OSTP) in having this conversation now: the pace of AI development and deployment is so rapid that favorable and unfavorable world-changing consequences are possible in a brief time. In the following pages, we draw on the broad and deep expertise of Mathematica staff across disciplines and policy domains to address many of the critical questions posed by OSTP to inform a National Artificial Intelligence Strategy. Our own use of AI across contexts—from public health to education to employment and labor—to inform public decision making informs our responses to OSTP's questions. For example, our previous work has featured several relevant skills:

- We explored the value of wastewater data to anticipate surges in communicative diseases.
- We used machine-learning techniques to identify academically at-risk students.
- We have designed and trained AI solutions for a challenge sponsored by the Centers for Medicare & Medicaid Services (CMS).
- We predicted fatal opioid overdoses using geospatial analytics.
- We employed agent-based models to predict the spread of COVID-19 in schools under various operational conditions.

In these and other projects, we have shown how AI can be a powerful tool to support decisions and augment services that promote public well-being in general and equity in particular. Our experience with AI tools has also informed us about many of the associated risks, including the ways that poorly designed or implemented AI systems might increase inequities. This is a particular challenge for predictive analytic systems used for making high-stakes decisions (such as personnel evaluations or criminal sentencing); for example, AI systems can perpetuate existing racial biases even if they are nominally race-neutral, because they learn from historical data containing implicit racial biases or reenforce racially biased decisions. Just like any other decision-making tool, how AI is deployed can promote either equitable or inequitable outcomes.

Our response is informed by more than a half century of rigorous analysis of organizational systems for implementing policy and delivering services, in domains such as health, education, employment, or international development. The effects of AI on these systems are likely to vary substantially, depending not only on the particular AI tools in question but on the differing institutional structures and features. Mathematica knows these institutional features well; therefore, we can anticipate some of the ways AI's impact might differ across sectors.

Item-by-Item Response

Protecting rights, safety, and national security

1. **What specific measures—such as standards, regulations, investments, and improved trust and safety practices—are needed to ensure that AI systems are designed, developed, and deployed in a manner that protects people's rights and safety? Which specific entities should develop and implement these measures?**

Measures taken to protect rights and safety must account for the fact that several different types of AI systems exist, and AI systems evolve over time. Over the past decade, we have seen an increase in the use of AI systems to guide decision making and resource allocation in sectors where protecting people's rights and safety are of the utmost importance, such as health care, education, transport, employment, and welfare. To date, most of these AI systems have used classification and ranking algorithms for which research and frameworks exist to support responsible design, development, and deployment. In contrast, far less guidance is available on how to responsibly build and deploy AI systems that use performant large language models (LLMs) such as GPT4, which are being deployed rapidly to execute various tasks, including customer support, content creation, and translation. The pervasiveness of AI begs the question: Is AI now a social determinant of health and well-being?

Looking to the future, people's rights and safety must consider even broader applications of AI that do not yet exist yet: artificial general intelligence (AGI) (Stuart Russell 2019). Although there is no consensus about when AGI will arrive, the possibility of a misalignment of interests between humanity and a super-intelligent AGI is real and demands the attention of policymakers well before the creation and deployment of AGI. The possibilities of AGIs motivate a need for investing in AI alignment research, with the aim of ensuring safe solutions are built in to future AGIs when they launch.

Appropriate regulations require improved understanding of AI's risks and safety on the part of policymakers and the public. Philanthropies and government agencies should therefore invest in translational media to inform decision makers and the public about AI's risks and safety. Increased collaboration between academic researchers, policymakers, and the technology sector could promote translational work. Translational media (such as the kinds of infographics, blog posts, podcasts, webinars, and conferences that Mathematica frequently supports) can play a key role in elucidating what AI systems look like in practice, providing policymakers with enough understanding so they can collaborate with researchers and technologists to develop feasible standards and relevant regulations.

In the case of LLMs, the most pressing need is research that establishes best practices for responsible design, development, and deployment. As shown by OpenAI's grant program for such research, there is a knowledge gap around how these systems have the potential to threaten people's rights and safety, and what best practices can be established to prevent or mitigate that threat. Although the 117th Congress proposed several bills that would regulate generative AI, filling the knowledge gap is key to ensuring these regulatory bills reflect the varied ways in which generative AI systems can threaten people's rights and safety. For example, some methods exist for detecting and measuring bias in generative AI outputs, such as adversarial training (that is, red teaming) and counterfactual methods, but there is a need for methods that quantify the extent of bias present in these models and the consequences of these biases (for example, how often does ChatGPT output harmful stereotypes and what are the impacts of those stereotypes if ChatGPT is used to power a

mental health chatbot?). Regulators should also consider the quality and representativeness of the data used to train LLMs.

2. **How can the principles and practices for identifying and mitigating risks from AI, as outlined in the Blueprint for an AI Bill of Rights and the AI Risk Management Framework, be leveraged most effectively to tackle harms posed by the development and use of specific types of AI systems, such as large language models?**

The Blueprint for an AI Bill of Rights and the AI Risk Management Framework play key roles in codifying guiding principles for responsible AI. Several responsible AI tools already exist that apply the Algorithmic Discrimination Protection principle of the Blueprint at each stage of the data science project life cycle (that is, model design, identifying and quantifying bias, mitigating bias, and explaining bias). Datasheets for Datasets or Carnegie Mellon University's Data Science Project Scoping Guide are useful during the model design phase to address potential ethical issues before they occur and ensure that the project is designed and implemented with equity as a priority. Once a model is built, several tools are available to assess fairness and quantify potential harms posed by the use of the AI system. Fairlearn and Aequitas are open-source toolkits that explain how a machine learning model can treat certain groups or individuals differently, especially vulnerable populations. If the fairness metrics identify bias, bias mitigation algorithms, such as those available in AI Fairness 360's open-source package, help to mitigate sources of bias and increase equity. Lastly, because it might not be possible to remove bias entirely, transparency and explainability tools such as SHAP and Model Cards are useful to disclose and describe any bias that remains in the model. These tools can help to put the Blueprint's principles into practice if they are embedded at each respective stage of the model development process.

Mathematica has experience with implementing principles into analytic best practices through our work developing decision-support algorithms across a variety of sectors, including health, education, and social services. At the start of every new project, Mathematica screens for potential high-risk or high-impact applications of machine learning that require further oversight and potential use of the tools and frameworks mentioned in the paragraph above. We offer two other examples of how to put the AI Bill of Rights and the AI Risk Management framework into operation:

The CMS AI Challenge asked all challenge participants to address model bias. Mathematica used mixed methods, such as co-designing the AI Solution with the Patient Advocate Foundation and adjusting models for the under-representation of Hispanic people in the training data. Also in the CMS AI Challenge, Mathematica used SHAP to help create an explainability dashboard that conveys model results to clinicians and patients.

In work to address inequities in contraceptive access, Mathematica uses agent-based modeling to simulate the effects of a contraceptive access program, capturing individual interactions and adjusting for inequities at the individual level, thereby mitigating potential harms posed by this AI system before they even arose.

3. **Are there forms of voluntary or mandatory oversight of AI systems that would help mitigate risk? Can inspiration be drawn from analogous or instructive models of risk management in other sectors, such as laws and policies that promote oversight through registration, incentives, certification, or licensing?**

Voluntary oversight of AI systems is not sufficient, given the collective action problem: individual firms know there is substantial private value in being first to develop new AI technologies, and

society bears much of the risk. This is a prototypical case for regulation by government—although exactly how government should regulate is a much harder question.

Some agencies, such as the Food and Drug Administration, have policies and procedures to regulate certain AI-enabled products, but most auditing of algorithmic bias to date is done reactively by academics and activists. These audits focus on widely used algorithms and share the results publicly, stimulating bias mitigation via public shame, lawsuits, or a subsequent government investigation. A public interest group conducted such an audit in 2016 of the COMPAS recidivism software; the audit revealed the tool's predictions were racially biased. This set off public debate regarding the predictive validity of AI-driven tools, especially those with proprietary algorithms like COMPAS, and whether it is ethical to use these tools in high-risk scenarios in which people's lives are severely and adversely affected if the model makes an incorrect prediction. Although these audits are high impact, government agencies must rely on mechanisms other than private actors to gather evidence needed to open an investigation into an algorithm—especially as AI systems continue to evolve.

One potential form of oversight might be to give these audits a more formal role, via a seat at the table within a government agency. The government could establish a bureau whose mission is to use audits as a means of enforcing legislation that regulates the use of AI. For this form of oversight to be feasible, the United States could draw upon legislation like the European Union (EU) AI Act, to establish concrete rules regarding the levels of risk posed by the different types of AI systems and the allowability of such systems based on their level of risk. In the early stages, the government could contract with a third-party organization to conduct these audits and then build in-house capabilities over time.

In addition to downstream audits after AI systems exist, another approach is to focus on upstream risk management, such as the responsible AI tools we discussed in response to Question 2. For example, regulators could push for more public assessments of data quality, such as the work Mathematica conducts on behalf of CMS to assess the Transformed Medicaid Statistical Information System (T-MSIS) data. Next, the government could encourage the development and use of algorithmic fairness tools, such as Mathematica's Bayesed and Confused, an open-source application to quantify algorithmic fairness. In addition, regulators could encourage the use of Bayesian-backed AI systems to quantify the uncertainty when assessing model risk; Mathematica demonstrated the Bayesian approach's utility when we used it to stabilize measures of school performance, in turn reducing the likelihood that policymakers will make consequential decisions based on random error.

Advancing equity and strengthening civil rights

9. What are the opportunities for AI to enhance equity and how can these be fostered? For example, what are the potential benefits for AI in enabling broadened prosperity, expanding economic and educational opportunity, increasing access to services, and advancing civil rights?

For AI to enhance equity, AI-driven solutions must effectively address the problems they seek to solve. As of today, few AI solutions are evaluated in a rigorous way to determine effectiveness. One example of a well-evaluated AI platform is Bayesian Health's clinical platform backed by research showing reductions in mortality. Researchers and regulators should promote frameworks on how to evaluate the effectiveness of AI solutions, such as rapid-cycle evaluations, because large experimental designs are not always possible—especially for large-scale, consumer-facing AI tools.

When effectiveness is determined, fostering the benefits of AI-driven solutions depends on affordability, acceptability, and accessibility, with sustained maintenance and improvement over time. For example, intensive human tutoring is known to be academically effective, but it is expensive and difficult to access due to limited availability of tutors. AI tutoring, if effective, could address educational gaps if it is used, accessible, and affordable to students; however, because curricula and knowledge change over time, the AI solution must continue to update and improve over time. AI systems have the potential to deteriorate (that is, model drift) over time, and any inequities in how the model treats individuals on the basis of protected classes often deepen when this deterioration occurs. Therefore, model maintenance and updates over time are crucial to ensuring AI enhances equity rather than inequity.

Using health care as an example, Mathematica used AI to predict unplanned hospital admissions and mortality for the CMSAI Health Outcomes Challenge. Doctors can use these predictions to inform decision making, in turn increasing access to services. However, to ensure this happens in an equitable way, Mathematica considered affordability, acceptability, and accessibility in our submission. Working with our clinical and patient advocate partners, we used a human-centered design approach to develop the model and address concerns over its utility and accessibility, because doctors were concerned about yet another point-and-click solution. We drew on open-source data tools from federal agencies (namely the Agency for Healthcare Research and Quality and CMS) in an attempt to reduce the cost of ongoing maintenance and align with definitions used in industry.

Meanwhile, in K–12 schooling, Mathematica (with the support of the U.S. Department of Education's Mid-Atlantic Regional Educational Laboratory) has applied state-of-the-art Bayesian statistical methods to improve the accuracy of measures of the performance of historically disadvantaged groups of students (Forrow, Starling, and Gill 2023). This is an excellent example of how advanced analytic techniques can promote equity by ensuring the performance of all student groups is both visible and accurate.

10. What are the unique considerations for understanding the impacts of AI systems on underserved communities and particular groups, such as minors and people with disabilities? Are there additional considerations and safeguards that are important for preventing barriers to using these systems and protecting the rights and safety of these groups?

As discussed in our response to Question 9, the nation needs a framework to evaluate the effectiveness and impacts of AI across a broad set of factors. In addition to addressing the target issue, an interdisciplinary approach must consider other factors—not just the predictive accuracy of the model or the ability to improve outcomes overall. Because minors and people with disabilities are a subset of the full population, evaluations should be designed to assess impacts for key subgroups.

One idea is to draw upon the risk framework set forth by the EU's AI Act, which provides useful considerations for AI uses and risk factors, with specific definitions for vulnerable populations. The EU AI Act categorizes potential risks by four levels: minimal risk, limited risk, high risk, and an unacceptable level in which AI is too risky to use. Mathematica's work and expertise is predominantly in contexts in which applying AI would be considered high risk, given that most of our projects are in the health, education, employment, and social services sectors, and all AI systems in these sectors are classified as high risk. To assess these risks, Mathematica looks at the outcomes in a specific population, retrains and validates the model for a population, or applies guidelines or safeguards to prevent the application of models to certain populations.

One specific area to focus on is AI to support decisions based upon assessments for high-risk individuals undergoing a transition. For example, a higher-risk field that currently uses AI in society is providing reentry services to formerly incarcerated individuals. In a Mathematica survey of U.S. Department of Labor (DOL) Reentry Project grantees, 84 percent reported using participants' risk and needs assessments for screening or service planning, and most grantees used more than one assessment (Stein et al. 2023). Using risk and needs assessments for service planning and program screening can help staff develop targeted treatment plans and screen out participants who might not benefit from services (Duran et al. 2013; Taxman and Smith 2021). However, potential issues associated with risk and needs assessments and the associated AI tools to support decision making could compromise their usefulness or introduce harms to participants. Predictive validity statistics alone are insufficient, as they depend on a sample of data that might not represent the intended population or might not represent how the AI system is used in practice.

Potential strategies to mitigate these issues are for agencies to use a risk and needs assessment that an independent third party has evaluated and tested with the populations they serve (Desmarais et al. 2022; Electronic Privacy Information Center 2020; Larson et al. 2016), as well as plan fidelity reviews to monitor implementation of risk and needs assessments (Wormith and Bonta 2018[1]). This Mathematica brief outlines these equity-driven challenges to using risk and needs assessments, and potential solutions in more detail.

12. What additional considerations or measures are needed to assure that AI mitigates algorithmic discrimination, advances equal opportunity, and promotes positive outcomes for all, especially when developed and used in specific domains (e.g., in health and human services, in hiring and employment practices, in transportation)?

A major consideration in reducing algorithmic discrimination is to clearly define who is accountable and responsible for safeguards throughout the AI life cycle. In Mathematica's experience, no single team or government agency can achieve this alone, and therefore the work must be interdisciplinary and involve public–private partnerships. Similar to what some federal initiatives that Mathematica has participated in now use, federal agencies should consider a standing socio-technical working group (STWG) and a field working group (FWG). The STWG represents a broad set of perspectives to develop frameworks to support regulation, whereas an FWG would work to evaluate existing applications to identify risks. Depending on the scope of the AI, these interdisciplinary teams would include cultural and economic sociologists, technical experts, policy analysts, legal experts, and social workers (the STWG) as well as organizations and individuals who would be affected by the technology, such as community-based organizations, civic and religious leaders, educators, social workers, and community members interested in AI (the FWG). We see opportunities for adapting policy analysis techniques like rapid cycle evaluation and Learn, Innovate, Improve (LI[2]) that can mitigate algorithmic discrimination and promote public well-being.

This kind of interdisciplinary review of AI applications is neither ancillary nor optional. For larger applications of AI, regulators could consider this a required part of the product development process, just like prototyping or initial design for a product, or homologation for an automobile. To encourage innovation, the federal government could design structures that facilitate access to this kind of mitigation review for organizations without the bandwidth to implement this on their own.

[1] Wormith, J. S., and J. Bonta. "The Level of Service (LS) Instruments." In Handbook of Recidivism Risk/Needs Assessment Tools, edited by J. P. Singh, D. G. Kroner, J. S. Wormith, S. L. Desmarais, and Z. Hamilton (pp. 117–145). John Wiley & Sons, 2018.

Bolstering democracy and civic participation

14. How can AI be used to strengthen civic engagement and improve interactions between people and their government?

One recent technology to examine to inform AI regulations is the use of blockchain technology. On the most basic level, blockchain technology is used as a so-called smart contract to facilitate secure transactions; AI-driven applications could, if designed correctly, facilitate smart transactions between people and their government. For example, applications for Social Security Disability Insurance (SSDI) are time intensive, require many inputs, and it can be frustrating for the applicant to understand where they are in the process. Paperwork reduction and document receipt confirmation are key challenges for programs like SSDI. An AI-enabled application built on a purpose-built blockchain could facilitate the secure and transparent population of SSDI applications across relevant parties (doctors, government, applicants), and the use of such a tool could facilitate trust between the parties that information is accurate, complete, and timely.

Implemented correctly, AI-driven solutions for government interactions could also drive engagements or improved interactions through scale but must support multiple modes and platforms. Public services have variable demand, and AI-driven solutions could respond to inquiries or send nudges. Examples include deploying AI bots to steer citizens to useful resources or send reminders on public common platforms (like social media) or connecting inquirers to on-demand services (like rideshares). These structures offer government the opportunity to create public digital spaces and platforms that could facilitate service delivery, constituent engagement, and civic engagement, with an eye towards innovating a 21st century digital public commons.

While AI infused smart contracts offer opportunities to expand public well-being, they also present significant risks. Within most blockchains, there is no one party responsible for its contents—or its compliance with regulations. This ability of AI to respond to needs must be counterbalanced with concerns around security and surveillance. The integration of AI into blockchain will raise the stakes of sensible and right-sized regulation of both AI and blockchain.

Promoting economic growth and good jobs

19. What specific measures—such as sector-specific policies, standards, and regulations—are needed to promote innovation, economic growth, competition, job creation, and a beneficial integration of advanced AI systems into everyday life for all Americans? Which specific entities should develop and implement these measures?

According to AI expert Suchi Saria, the most effective AI is developed for domain-specific use cases. Thus, although there is a need for clear and transparent cross-cutting regulation, sector-specific regulations are important to meet the domain-specific AI use cases.

One broader example of AI standards involves the hiring process. Given the state of AI and human resource technology today, one could imagine an entirely AI-enabled hiring process from the applicant's and the employer's perspective—an automated generation of a resume and cover letter feeding into an automated applicant review and scoring process. The government could provide specific guidance on discriminatory practices or create base technology to make these AI applications more accessible to all job seekers and employers (if desirable).

On a more specific use-case by use-case basis, a major measure to consider is monitoring compliance; thus, enforceability is a major concern. One idea is to create a certification program,

similar to how electronic medical record software is regulated. Examples of how to monitor for compliance for specific industries and populations could look at Mathematica's work with DOL's Office of Federal Contract Compliance (OFCCP) to examine the implementation and effects of new training on OFCCP's mission of compliance. In addition, monitoring could examine how agencies support employers to comply with AI standards (see Mathematica's evaluation of DOL's effectiveness of bringing employers into compliance with labor standards). New standards surrounding AI will call for the same type of rigorous evaluation and continuous improvement. As new policies and standards around AI emerge, both effective training on new compliance standards and evaluations of the strategies used to enforce those rules and regulations will be key ingredients to an effective government response.

22. **What new job opportunities will AI create? What measures should be taken to strengthen the AI workforce, to ensure that Americans from all backgrounds and regions have opportunities to pursue careers in AI, and otherwise to prepare American workers for jobs augmented or affected by AI?**

The conversation around AI and jobs often focuses on job opportunities that could be lost due to technology, but there is also opportunity to create or improve jobs through AI. We suspect this opportunity brings risks that the American worker is not prepared for, and we see policy as having a role in facilitating the job-to-job transitions that will enable AI to achieve the goal of generating opportunity.

To ensure that Americans, regardless of background, can gain the skills needed to thrive in an economy influenced by AI, policymakers should ensure that resources are available to workers to make a variety of career paths are viable. This includes preparing for the new economy at the early stages of education and for adult workers who might want to transition to new work. Curriculums are likely to continue to include more topics surrounding both technical and complimentary skills; skills such as critical thinking and strong communication have become increasingly important for hybrid jobs that combine technical skills with complimentary soft skills. For workers directly affected by AI or automation, policies such as retraining assistance or temporary income support are worth exploring.

New curriculums and training programs will need iteration as AI continues to evolve. Assessing the efficacy of our education and training will enable the focusing of resources into those programs shown to be both effective and equitable. Mathematica has experience in considering both these attributes in projects like the Middle Year's Math Evaluation for adult education and training programs Mathematica conducted for an evaluation called Assessing Evidence of Effectiveness in Adult Education for the U.S. Department of Education to best identify best practices, approaches, and strategies in adult education that could be applied to new curricula during job transitions.

Innovating in public services

24. **How can the Federal Government effectively and responsibly leverage AI to improve Federal services and missions? What are the highest priority and most cost-effective ways to do so?**

The federal government can leverage AI across programs and services to effectively carry out their missions. AI could reduce the administrative budget for citizens and government employees, add new features to programs, or assist decision making. The overarching principal should be to use AI that is designed and evaluated for the specific use case to address major barriers to services.

Building on experiences in the opioid epidemic and COVID-19 pandemic, government at all levels should consider how AI tools can assist at each stage of a disaster or emergency. We see the possibility of such tools as especially strong in the context of expanding incidents, in which the nature of a disaster is unclear or changing and the level of impact appears to increase. Our recent report on human services and disasters showed that a lack of good data—namely, where those displaced by a disaster headed and the human services supports needed by the receiving locales—hampered disaster response and early recovery efforts. By the time these gaps rise to the level of government actors, the resources of local community-based organizations are stretched. Using predictive analytics and related tools, we believe emergency managers can more proactively position needed resources. Sun and colleagues (2020) gave specific suggestions on the types of AI tools that might be appropriate at each stage of the disaster life cycle.

As we saw in our work on pandemics and wastewater monitoring, AI tools are most effective when the appropriate data sets and corpora are appropriately integrated into the analytics. To track pandemics, it seems obvious that detecting the pathogen (the virus that causes COVID-19, for example) was crucial. What we learned through our work is that linking the emergency to its cause is not as simple as linking a pandemic to wastewater: we must situate that causal component within a larger context of local public health policies, community demographics, and social vulnerability. That is where truly useful insights emerged, and it is where the most vigilance is needed to ensure AI is deployed legally and equitably.

Crucially, deploying AI tools in the disaster life cycle will *not* eliminate the reproduction of mitigation, response, and recovery inequalities that often occurs, with disparate impacts often playing out on factors like geography, socioeconomic status, homeownership, and race. AI deployed within the disaster life cycle will offer rapid suggestions for emergency managers and the voluntary sector (VOAD). The organizations who use these tools must collaborate with the affected communities and techno-sociologists who can identify where AI might reproduce inequalities and inequities.

About Mathematica

Mathematica is a research and data analytics consultancy of nearly 2,000 professionals driven by a mission to improve well-being for people and communities around the world. We innovate at the intersection of data science, social science, and technology to translate big questions into deep insights for public and private sector partners that count on our expertise and our commitment to equity. Collaborating closely with a wide range of decision makers and changemakers in health, human services, and international development, we reimagine how the world collects, analyzes, and applies data to solve urgent global challenges. We center our work on diversity, equity, and inclusion, with equity-informed and culturally responsive practices fundamental to the professional services we offer, including research and evaluation, data collection and analytics, technical and implementation assistance, and evidence-backed support for decision makers.