MERVE HICKOK AND EVANNA HU

Don't Let Governments Buy AI Systems That Ignore Human Rights

Even in the absence of broader AI regulation, federal procurement provisions could set expectations for data quality, model performance, risk assessments, and documentation.

In 2022, 28-year-old Randal Reid was driving to his mother's house for Thanksgiving dinner in DeKalb County, Georgia, when he was pulled over by local police and arrested for a crime committed three states away in Jefferson Parish, Louisiana. But Reid had never been to Louisiana, let alone Jefferson County. The sheriff's office there had taken surveillance footage showing a Black man stealing designer purses and fed it to commercial facial recognition software. The software misidentified Reid, who is also a Black man, and led to his arrest. It was not until December 1, after Reid spent a week in jail and thousands of dollars on legal and other fees, that the Louisiana department acknowledged the error and Reid was let go.

Wrongful incarceration is one of several types of human rights violations that unregulated and irresponsible use of AI systems may lead to, but local, state, and federal government procurement regulations in the United States do not require vendors bidding for government contracts to conduct assessments for the quality of data used to build their products, or for their products' potential bias, risk, and impact. Facial recognition software—including programs sold by IBM, Amazon, and Microsoft—has demonstrated accuracy of more than 90% overall, but a landmark 2018 study found error rates can be more than 30% higher for darker-skinned women than for lighter-skinned men. The population of Jefferson Parish is 36% non-white, but, to our knowledge, the sheriff's office procured the system from Clearview AI without performing any specific assessments of its performance or potential for social harm. *(The office did not reply to* Issues' *emails asking how it assessed the software.)*

In the United States, most AI systems are procured by federal, state, or local entities the same way as traditional software; criteria focus on the cost of a project and a vendor's past performance. Vendors are not required to, say, demonstrate that their solution can perform as needed in the real world. Systems developers need not prove the provenance or quality of their training data, share their models' logic or performance metrics, or lay out design decisions, such as the trade-offs they made and the risks they foresaw and accepted. (Where guidelines do exist, they are outdated and limited to assessing risk for privacy and cybersecurity; one current Department of Justice guide dates to 2012.) Vendors can even claim trade secrecy to deny requests for such information, foreclosing critical assessment and independent validation.

Meanwhile, AI tools are increasingly used for applications that could upend people's lives: criminal investigations, housing placements, social welfare screening, school assessments, and felony sentencing. In 2019, the story broke that for six years, Dutch tax authorities had been using a self-learning algorithm to help create risk profiles for identifying childcare benefits fraud. Largely based on the algorithm's risk indicators, authorities wrongly fined and demanded tax repayments from tens of thousands of families—who tended to be from ethnic minorities or have lower incomes—and pulled more than a thousand children from their families into foster care. Families were impoverished and destroyed; some victims committed suicide. In 2021, the prime minister of the Netherlands and his entire cabinet resigned over the scandal.

There are other examples of poorly wielded AI technologies that cause life-altering harm. In 2016, the United Kingdom relied on automated voice analyses to verify the identities of those taking English language tests, resulting in around 7,000 foreign students being falsely accused of cheating. The students' visas were revoked, and they were asked to leave the country. In 2022, US Customs and Border Protection deployed biased facial recognition software that did not accurately detect Black faces, effectively blocking applications from many Black asylum seekers. More and more research papers and news accounts reveal problems resulting from poor training data, unrecognized misassumptions, and misuse of tools intended for good.

The advance of artificial intelligence has been accompanied by a steady stream of high-level recommendations on how to regulate the technology, including the creation of new agencies. Blue-ribbon groups have worked out AI principles or frameworks that focus on the development or governance of AI systems. The White House published the Blueprint for an AI Bill of Rights in 2022; President Biden issued an executive order in 2023 laying out overarching standards to protect Americans' privacy and civil rights that lists more than a page of AI uses presumed to impact rights. But there is a chasm between making a smart list and instituting real protections—one that can only be bridged through concrete steps. These include legislative action from Congress to regulate private actors as well as formal guidance from the Office of Management and Budget (OMB) on federal use of AI (all of which have yet to materialize). In the meantime, the US government is embedding AI systems into its infrastructure without robust and consistent safeguards.

Federal procurement policy could quickly put this in check by demanding that all federally purchased AI systems respect human rights. This would not only prevent dangerous purchases within the federal system, but could also become a model for state and local governments as well as other entities. Although robust regulations are unquestionably important, mandatory mechanisms embedded within the federal procurement process could go a long way to enhance accountability and avoid societal harm, both before and after other regulations are in place.

The power of procurement

For years, we have been advocating for more accountability for public actors and firmer centering of human rights in how governments procure and use AI. Hickok has testified in Congress, assessed dozens of other governments' use of AI systems, researched public procurement of AI, and submitted recommendations to the federal government. Hu was a consultant on the Department of Defense's Ethical Principles for Artificial Intelligence and has briefed dozens of policymakers and senior executives on trusted and secure networks and supply chains and adaptable procurement processes designed to acquire emerging technologies, such as Other Transaction Authority.

We think procurement guidance can be a particularly effective regulation tool because government agencies rarely develop AI systems de novo—they procure commercial ones or hire services from AI vendors. Considering civilian federal agencies alone, there are now more than 1,200 already deployed or planned use cases for AI, according to the US Government Accountability Office. Non-defense agencies requested $1.8 billion in the 2023 federal budget to implement AI technologies. (The Department of Defense itself received $1.8 billion, and has disclosed some 700 ongoing AI projects.)

Instituting federal procurement guidance at this stage would have cascading effects across the AI marketplace. State and local funding for AI tools often comes from federal agencies, such as the Department of Homeland Security or the Department of Justice (via the National Institute of Justice grants), which require compliance with federal procurement guidelines. And when states and municipalities do make their own purchases, they refer to federal guidelines to establish their own practices. Moreover, most AI vendors marketing to state and local clients also have contracts with federal agencies and so create tools with national regulations in mind. Plus, the US federal government is the largest buyer in the world and seen as a role model by other countries. Even in the absence of broader AI regulations, specific procurement provisions could set expectations for what AI vendors should provide in terms of data quality, model performance, risk assessments, and documentation.

For years, we have been advocating for more accountability for public actors and firmer centering of human rights in how governments procure and use AI.

To serve the people

Federal procurement could also do much to shift which groups AI systems are built to serve. Today's incentives encourage the design of tools for government employees, not for those people most affected by algorithmic recommendations. After relevant personnel within a government agency decide what problems to solve and provide technical requirements, dedicated procurement officers oversee the purchase and ensure compliance with federal regulation as well as any applicable international regulations. But there is a disconnect: those selecting what to purchase may never interact with it again. Intended "end-users" might be social workers, immigration officers, soldiers, or recruiters. Procurement guidelines could help ensure that the needs of multiple people are considered in the process.

In particular, procurement offers a chance to represent the interests of those directly affected by decisions made with AI tools—Randal Reid, for instance, and child welfare applicants, job candidates, and refugees awaiting asylum decisions, to name a few. We think it is remarkable that today's discussions of AI have no uniform language for the category of people with the most at stake. International development non-governmental organizations call them "beneficiaries." The Australian government calls them "customers." The European Union AI Act calls them "individuals" or "impacted persons." The Council of Europe's draft convention on AI mentions "persons interacting with AI," "affected persons," and "persons concerned." OMB's draft guidelines on federal use of AI calls them "impacted individuals" or "customers" but those terms can concurrently mean "individuals, businesses, or organizations that interact with an agency."

The very fact that there is no uniform term for the group of people most affected by AI tools shows how little their rights are considered in any stage of AI production and deployment—and how poorly prepared industry and governments are to take their needs into account. Companies have plenty of experience designing systems centered on their imagined end user, but we think governments should only purchase systems that demonstrably center the human rights of persons likely to be impacted by them. Procurement standards would ensure that governments uphold their chief duty: to serve the people.

Governments may be able to choose among bids from multiple AI vendors when making a purchase, but they have a monopoly on public services like policing, health care, and public welfare. In the case of Randal Reid, it was not the purchaser or end user of the facial recognition software who was forced to spend a week in jail; it was someone with no say in the system at all. (Indeed, Reid needed legal assistance to even find out that an AI system had identified him, presumably from photos he'd posted on social media.) Often, the more vulnerable the individual, the more they must rely on public services—and the more they are subject to enforcement. Thus, people with the least power are those most exposed to decisions made with AI systems. AI tools are widening the digital divide around public services because the most vulnerable are also less likely to have internet access or other resources (education, connections, implicit knowledge) needed to, say, opt out of default data collection. Whatever people subject to decisions made with AI are called, they deserve fair treatment and recourse to justice.

A procurement framework based on human rights

Relying on definitions in the United Nations' Universal Declaration of Human Rights, it's clear that AI systems impact such fundamental rights as the right to privacy, equal protection against any discrimination, access to social security, and access to effective remedies. AI systems can also impede freedoms, such as freedom of expression, association, or freedom from arbitrary arrest.

We think building protection of human rights into procurement decisions (as well as AI design) requires several steps. First is justification that an AI system is indeed a solution. This entails collecting evidence demonstrating why it performs better than other methods and that it can actually work as intended within its operational context. The second step is to formally assess assumptions and design decisions within the AI system with an eye to gauging positive and negative impacts on communities. The Office of the Director of National Intelligence, for example, has published a six-page guide for evaluating an AI system's appropriateness and potential flaws. The third step depends on that kind of assessment. If it foresees harms, measures should be taken to mitigate them. If mitigation is not possible, the AI system should not be procured. The fourth step is to ensure that the system is transparent enough to allow contestation and legal challenges.

Here are broad requirements for a procurement process that centers human rights.

Prohibit AI systems based on scientifically invalid premises. Too many AI systems are designed around spurious concepts and correlations. Many prevalent approaches, including biometric categorization, biometric emotion analysis, and predictive policing, lack scientific validity. Though vendors claim their systems are objective and capable of predicting or inferring such things as a person's emotions, ethnicity, or likelihood of committing a crime, consensus is growing that they instead identify spurious correlations between disparate data points. The EU AI Act prohibited the use of these systems in many high-stakes contexts. The UN High Commissioner, European Data Protection Board, and European Data Protection Supervisor similarly called for a ban on the use of AI systems in public spaces that identify individuals (such as facial recognition technology), saying, "While the justifications for such programmes are currently theoretical and lack supportive evidence, the harms have been real and, often, irreparable." More than 2,000 researchers signed a letter condemning crime prediction technology based on biometric and criminal justice statistics. Even in the United States, the Government Accountability Office recommended restricting funding to the Transportation Security Administration's behavioral risk assessment system, which lacked scientific validity.

We would like to see US procurement guidance require vendors to back their claims with peer-reviewed research. Procurement teams cannot be expected to have the expertise to evaluate vendors' claims on their own.

Enable meaningful transparency. Similar to the US Food and Drug Administration's requirement for nutrition labels that clearly describe macronutrients, micronutrients, and allergens, AI systems and outcomes need to be more transparent about listing key characteristics. Proposals for how to do so have already been put forward by academics and implemented by business communication company Twilio. These kinds of disclosures would not require vendors to hand over their source code or even ontology, but there is other essential information—the parameters of their training data, rationale behind the chosen model and performance metrics, methodology of their testing, and outcomes of such testing—that should be available for public inquiry. That kind of transparency fulfills requirements that agencies notify the public and collect input before advancing certain programs, allowing others to assess the vendor's design decisions, the fitness of the AI system for the purpose or context, and the policy choices embedded into the AI system. Importantly, to bring accountability, transparency criteria must be fulfilled before a contract is awarded. If transparency and external stakeholder engagement occurs only after procurement, it might be hard to terminate the system and the contract.

Robust transparency practices and expectations would powerfully aid advocacy on behalf of human rights. Academic and journalistic engagement have led to effective monitoring across industries. Efforts like the AI Incident Database and AI, Algorithmic, and Automation Incidents and Controversies Repository do much toward opening these systems to inquiry, but cases are opened only when problems are reported, and essential information often goes missing. Procurement can push to broaden transparency with AI systems that are explainable, traceable, contestable, and subject to third-party testing and verification. And all of this information can be brought into AI registries, enabling more advocacy with concomitant improvements in practice.

> Governments should only purchase systems that demonstrably center the human rights of persons likely to be impacted by them.

Conduct human rights assessments. Any major procurement process should have an obligatory assessment of human rights implications. It should cover three major aspects: vendors' own analysis, agencies' domain expertise, and meaningful engagement from external stakeholders, such as the broader public and civil society organizations. That last task may sound chaotic, but the United States has precedent and best practices to get public comment for policy changes that could apply to high-stakes AI procurements. A comprehensive and easy-to-use reference for assessing human rights in this context is the Dutch Ministry of the Interior's Fundamental Rights and Algorithm Impact Assessment, which splits the assessment into four themes: intended effects (objective), data (input), implementation and algorithm (output), and fundamental rights (impact). Practically, documentation would be collected similarly to existing protocols, such as the required sections on monitoring and evaluation in State

Department and US Agency for International Development requests for proposals, or the Cybersecurity Maturity Model Certification requirement for all Department of Defense vendors. When vendors' proposals are evaluated, human rights assessment should rank as highly as other major criteria, such as desired outputs and key metrics of effectiveness.

Use ongoing assessment to ensure public infrastructure systems are adaptable. Once a technology is deployed, it blends into the background, and its assumptions are not questioned again. The focus of the agency shifts to system maintenance, rather than considering what kind of policies and values the technology promotes or what effects it is having on communities. As law, technology, and society scholar Andrew Selbst and colleagues put it, "code calcifies," locking in any of several possible conceptual traps whereby technical systems fail within social worlds. The emphasis on maintenance preserves the status quo and deprioritizes responses to changing circumstance or inequities. But AI procurement processes and contractual clauses should allow for flexibility in AI approaches because systems and models are still evolving. The research and product landscape shifts continuously. Just as procurement provisions should build in processes for transparency, so too should they secure ongoing assessment of the system and its assumptions, perhaps tied to contract renewals or retained services.

Time to set requirements

The Office of Management and Budget is expected to finalize guidelines on federal use of AI later this year. It was two years late getting started on its mandate to draft them, but as part of the public consulting process, it collected a wealth of recommendations on how federal agencies should develop, design, procure, and use AI systems.

We urge OMB to finalize its guidance without further delay and leave no ambiguity on the rigorous controls that should be implemented: bright-line prohibitions against unscientific AI systems, clear mandates for evidence-based justification, and impact assessments on human rights that are contestable, explainable, and traceable. Furthermore, we think that OMB should be stricter than current draft guidance about requiring assessments and controls to take place before an AI system is put in use. Assessment should happen *during* the procurement process, when work is still being specified and the purchaser has more leverage—not as an afterthought or after the contract is awarded.

Critics will counter that strict procurement requirements will increase complexity of an already burdensome process and could drive private companies away from public contracts. But it's hard to imagine successful businesses leaving such a big market unserved. Another objection is that US human rights requirements could increase the costs of systems or the alacrity with which they are adopted and so give China a competitive edge selling AI technology to less conscientious governments. These arguments are analogous to suggesting that passenger planes should be designed without seats or seatbelts or pressurized air because it is simpler and more efficient. The function of democratic governments is protecting their citizens. As the US Department of Justice, Consumer Financial Protection Bureau, Federal Trade Commission, and Equal Employment Opportunity Commission jointly stated, vendors should be held accountable when they fail to address discriminatory outcomes. And vendors can already move toward fulfilling these kinds of requirements and begin to, as a UN group developing guidelines has stated, "operationalize respect for human rights as part of how they do business."

> We would like to see US procurement guidance require vendors to back their claims with peer-reviewed research.

In the absence of specific guidance, AI will continue to be treated as traditional software with minimum safeguards for human rights, which drastically underestimates its vast societal impact. Federal AI procurement will be piecemeal, incurring further risks to human rights as well as national security. Society should not depend on the personal initiative of procurement professionals, but instead be able to count on clear guidance and AI-specific procurement training.

Merve Hickok *is the president of the Center for AI and Digital Policy. She is also a lecturer at the University of Michigan and the founder of AIethicist.org.* ***Evanna Hu*** *is the CEO of Omelas, which tracks authoritarian propaganda, and a Forward Defense Senior Fellow at the Atlantic Council's Scowcroft Center for Strategy and Security.*