A PRAGMATIC, PARTICIPATORY, AND RIGHTS-PROTECTIVE MODEL FOR GOVERNING ARTIFICIAL INTELLIGENCE (AI)

For safe and ethical governance of AI, it is essential to create new and powerful mechanisms for community engagement and input, to protect basic rights, democratic values, and civil liberties, and to deepen investment in inclusive participation.

Advanced new AI technologies hold promise to expand opportunities, but also risk bias, unfairness, and opaque decision-making, and could perpetuate inequality at scale. Years of documented harms show that unchecked uses of decision algorithms lead to discrimination without accountability for those impacted. Our vision to address these serious concerns has three pillars—**Values, Voice, and Investment**. We propose them as an interrelated approach to AI governance:

- Honoring our shared **values** means acknowledging AI's current limitations and integrating strong ethical guardrails. This requires that we acknowledge and fully evaluate AI's risks and benefits compared to existing tools, and anchor uses to Constitutional principles and rights.
- **Voice** centers the need to create formal and powerful roles for impacted groups throughout the governance process.
- **Investment** would grow public capacity to participate in shaping technology and sharing in the prosperity it can bring to every community. This requires significantly expanding public interest capacity for impacts assessments, participatory engagement, skill-building, oversight, and to address worker impacts to democratize access to opportunity and ensure system design is fair and well-informed.

In its AI Executive Order, AI Bill of Rights, and the development of the Risk Management Framework, the Biden-Harris Administration is making significant strides on problem definition and principles. In our comments on the Office of Management and Budget's Memorandum on implementing the AI Executive Order, UnidosUS welcomed the draft guidance as a positive step towards ethical governance of AI.

The risks related to AI are substantial. Lending and credit access algorithmic discrimination persists despite legal authorities that require lenders to use the "least discriminatory alternative" and bar discrimination. As we saw with pulse oximeters during the pandemic, medical algorithms and design can also perpetuate health disparities if trained on unrepresentative datasets. To make matters worse, electoral misinformation targeting Latinos mischaracterizes the positions of political parties and candidates, and misinformation in Spanish is permitted to stay online far more than similar statements in English.

Latinos and other communities of color are also subjected to expansive governmental surveillance technologies. Predictive policing tools trained on flawed crime statistics have been found to disproportionately target low-income neighborhoods of color by falsely correlating race with criminality. Similarly, sentencing algorithms drawing on racially skewed conviction data likewise entrench harsher outcomes for minorities. Even in areas of strength for natural language AI systems, such as translation, because of the digital divide, much of the Internet (and thus, training data for models) is in English, and language barriers and differences in the quality of translations, may persist.
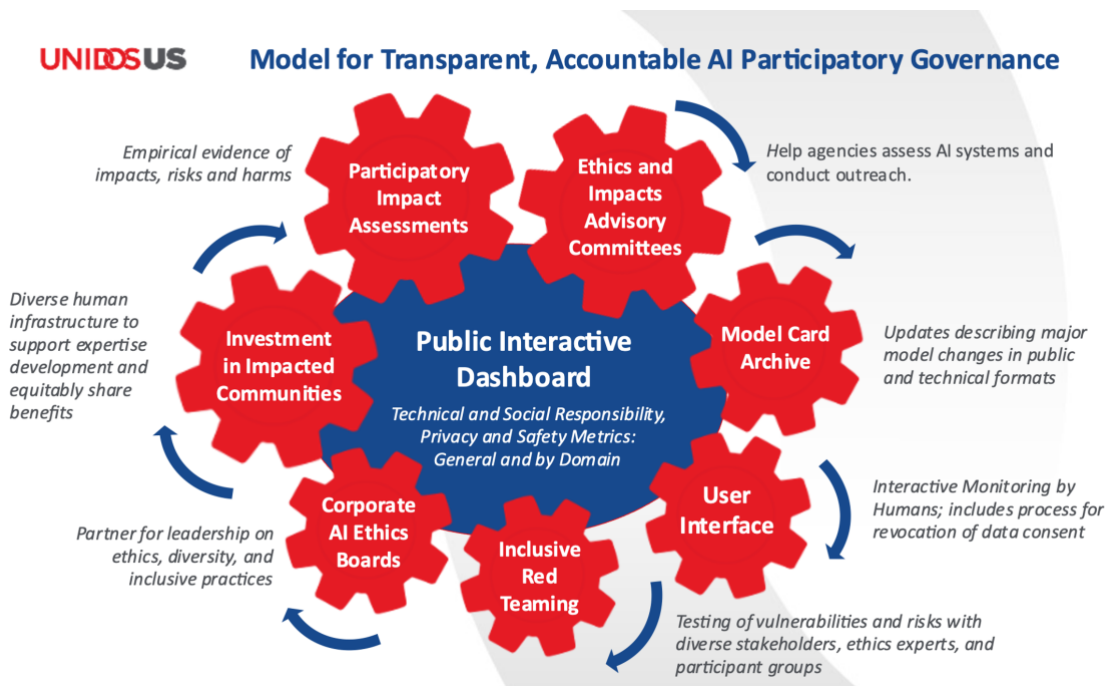
The powerful new AI capabilities and their rapid growth increase, the need for accountability, particularly when government power will back them. It will therefore be essential to create mechanisms

that democratize technology uses and provide practical ways for impacted communities and the public to weigh in on impacts in real-time feedback loops—we must move from technical assessments to socio-technical processes informed by democratic norms.

Democracies learn in public, and do not act until a deliberative process is completed that assesses harms and trade-offs, technical capacities and implications for shared values, and stakeholders can weigh in. AI governance, to be democratic in nature, should anticipate potential harms and include mechanisms for accountability to people they impact—including workers, creators, communities of color and lower-income workers, and others left behind and left out by traditional research, commercial incentives, and the digital divide (and who are thus invisible to the models). Too often, the bias or flaws in models are understood too late—so we must get better at both predicting and preventing foreseeable harms through good design: impacted groups are ideally positioned to tell technologists what they may not know they do not know.

Further integration of AI systems into government could eliminate these types of barriers, or it could create new, and potentially even more problematic, issues. Agencies should first audit existing algorithmic systems against principles of fairness, accountability, and transparency. Current or flawed tools should not simply be replaced with AI—agencies should thoroughly evaluate current AI uses and publicly describe their context and limitations before expanding adoption.

This mapping of current uses, alongside the creation of novel and strong oversight and inclusive participation in governance and norms-setting, will enable the learning needed to responsibly deploy new AI systems and guide constructive innovation. Linking ethical guidelines to participatory structures and capacity-building could, if done well, help to steer AI's benefits towards shared prosperity while earning public trust. Our comments propose a range of specific and practical mechanisms to address these issues, highlighted in the model below.

UNIDOS US   **Model for Transparent, Accountable AI Participatory Governance**

Empirical evidence of impacts, risks and harms — **Participatory Impact Assessments**

Ethics and Impacts Advisory Committees — *Help agencies assess AI systems and conduct outreach.*

Diverse human infrastructure to support expertise development and equitably share benefits — **Investment in Impacted Communities**

Public Interactive Dashboard
Technical and Social Responsibility, Privacy and Safety Metrics: General and by Domain

Model Card Archive — *Updates describing major model changes in public and technical formats*

Partner for leadership on ethics, diversity, and inclusive practices — **Corporate AI Ethics Boards**

Inclusive Red Teaming

User Interface — *Interactive Monitoring by Humans; includes process for revocation of data consent*

Testing of vulnerabilities and risks with diverse stakeholders, ethics experts, and participant groups

Voice—The Importance of Democratic Governance Structures for AI

- Strong oversight and community participation is needed for government to use AI ethically and maximize innovation potential.
- The task for the agencies should first be to thoroughly inventory uses, to create substantial new guardrails around *current* uses of AI tools in light of the NIST RMF, and to publicly identify the successes, caveats, criticisms from stakeholders, and shortcomings of these uses.
- Policy on current and new uses should be based on ethical guidelines linked to participatory design processes and expanded capacity, with multiple and overlapping opportunities for input.
- Impacted communities require formal structured roles and influence, not just 'check-the-box' perfunctory consultations.
  - Structures for participation should include AI Ethics and Impacts Advisory Committees with defined roles and input opportunities alongside public dashboards and user complaint mechanisms to monitor AI system performance informed by community feedback.
  - Continuous transparency mechanisms, such as a public dashboard, that publish indicators of capabilities, limitations and real-world impacts would improve transparency, drive accountability and productive innovation while educating the public.
  - Advisory Committees established at agencies can assess and catalog specific use-case AI risks, uses, and mitigate harm and assist with public outreach.
  - Impact assessments of use cases and development of an empirical body of evidence on the distributions of benefits and harms are needed to inform policy. Agencies should collect data on the experiences of affected populations to accurately describe how socio-technical systems operate in real-world conditions.
  - Inclusive red teaming exercises that stress-test AI systems are essential to uncover risks, biases, and failure modes pre-deployment. Intentionally integrating marginalized expertise helps uncover gaps technology teams may miss.

Values—Exempting AI Uses at the Heart of Constitutional Governance Would Undermine Democratic Norms and Incentives to Develop Technologies that Are Rights- and Privacy-Enhancing

- Many current government surveillance and data practices deeply concern individual privacy and civil liberties, especially for immigrant communities.
- Current government surveillance and uses of personal data frequently run roughshod over privacy rights and civil liberties, disproportionately harming immigrant communities.
- Although the NIST RMF framework calls for AI to be "privacy-enhancing," OMB's approach fails to ensure that this will matter where it is needed most. Instead, the Memo's proposed waivers could allow some of the most problematic and rights-infringing deployments of AI to continue to avoid even basic forms of public accountability and transparency.
- Safety- and rights-impacting uses, as in criminal justice and law enforcement functions essential to due process, should not be exempted from oversight.
- Exemptions would remove incentives to align systems with rights and values, yet effective law enforcement is compatible with sound and ethical technology design if privacy by design principles are applied.
- Hard cases cannot be the exception to our policies without undermining our fidelity to constitutional principles that rest at the core of our global leadership on personal freedoms and as a beacon of democracy.

Investment—Growing an Ecosystem for Equitable Participation, Public Trust, and Innovation Insights

- Governing AI equitably requires investing in public interest capacity, and substantial public interest capacity-building is crucial for participatory oversight.
- Invest in efforts to close the digital divide, which renders diverse communities invisible online and in training data for AI models.
- Working with stakeholders, and through its budgetary function, OMB and agencies should develop and inform a process for grants directly to impacted communities and nonprofit public interest organizations focused on issues implicated in AI and ethics.
- Congress should establish a dedicated fund modeled on the CDC Foundation that could support digital skills education and training, community-based AI auditors, participatory technology workshops, and other capacity building to close knowledge and equity gaps.
- Partnerships, training programs and funding for community oversight can help democratize AI's opportunities, and efforts should expand access to digital skills building education, and skills-based hiring approaches.

Marginalized groups often bear the brunt of technological harms given gaps in access, digital skills, and representation. Their lived expertise is vital to strengthen oversight and alter the incentives for tech design to reflect democratic, rather than extractive, practices. Renewing American leadership on equitable technology will demand openness to new forms of collaboration and best practices. With enough intention, we can choose to govern AI and algorithms in ways that align with our values.

Ultimately, the choice that is often posed, for example, between data security and privacy, on the one hand, and effective law enforcement, on the other, is a false one—good privacy by design can make both a reality once appropriate incentives and protections are in place. And the same is true for other aspects of AI systems that undermine fairness and shared values—we can and must ensure that standards for new technology for use by our democratic government require that they be fair, transparent, accountable, and explainable and are developed with input from impacted communities.

Please contact Laura MacCleery, Senior Director, Policy and Advocacy at lmaccleery@unidosus.org or Claudia Ruiz, Senior Civil Rights Analyst, at cruiz@unidosus.org, with questions or for additional information.