

Final Reports from the Second Carnegie Ethics Fellows Cohort, 2025-2026

Water as a Right or Commodity?
Ethical Frameworks for Allocation
Decisions

A New National Security Architecture:
An Economic Security & Strategic
Investment Framework for the
Western Hemisphere

Redefining the Frontline: AI,
Healthcare, and Non-State Power

Their Data, Whose Decision? The
Ethics of Artificial Intelligence in
Humanitarian Response

**CARNEGIE
ETHICS
FELLOWS**

**CARNEGIE
COUNCIL** *for Ethics in
International Affairs*

The following project was produced by a working group from the second Carnegie Ethics Fellows cohort reflecting nearly two years of convenings, collaboration, and research. Each report in this special series examines a critical issue at the intersection of ethics and international affairs. The [Carnegie Ethics Fellowship](#) aims to develop the next generation of ethical leaders across business, government, academia, and non-governmental organizations.

Acknowledgments

Carnegie Council would like to acknowledge the below individuals for their mentorship of the Carnegie Ethics Fellows group research projects. Thank you!

Tanisha Fazal: Advisory Board Member, Next-Gen Leadership Initiative; University of Minnesota

Manuela Fernández Pinto: Advisory Board Member, Next-Gen Leadership Initiative;
Universidad de los Andes, Colombia

Christine Jakobson: Former Carnegie Ethics Fellow; BCG BrightHouse

Juris Pupcenoks: Advisory Board Member, Next-Gen Leadership Initiative; Marist College

Cover Image Credit: Adobe/Rabbit_1990

Contents

Water as a Right or Commodity? Ethical Frameworks for Allocation Decisions	2
A New National Security Architecture: An Economic Security & Strategic Investment Framework for the Western Hemisphere	16
Redefining the Frontline: AI, Healthcare, and Non-State Power.....	32
Their Data, Whose Decision? The Ethics of Artificial Intelligence in Humanitarian Response	50
Cohort Overview	64

Water as a Right or Commodity?

Ethical Frameworks for Allocation Decisions

NADAV AVIHAY

RUTH NASHIPAE MUIGAI

HARSH SURI



CREDIT: Adobe/Jaroslav Machacek

Summary

Water is essential for human life and global development. However, about 4 billion people face severe water shortages for at least one month each year, and 2.1 billion people still do not have safe drinking water (UN-Water, 2024; WHO/UNICEF, 2025).

Meanwhile, access to safe, affordable, and reliable water is increasingly threatened by the combined pressures of climate change, rapid urbanization, population growth, infrastructure underinvestment, the expanding demands of water-intensive industries, and most recently—data centers.

Data centers can use up to 300,000 gallons of water each day for mid-sized sites and up to 5 million gallons for large ones. By 2028, AI-related data centers in the United States could need up to 32 billion gallons each year (WRI, 2026). At the same time, the Environmental Protection Agency (EPA) estimates that 12.1 million to 19.2 million U.S. households cannot afford water services (EPA, 2024).

These pressures reveal a central ethical problem in contemporary water governance: How should societies allocate a finite and life-sustaining resource when commercial development, public welfare, ecological sustainability, and intergenerational equity come into conflict?

This report provides an ethical analysis of water allocation and governance in light of data centers' rapid expansion in the United States and globally.

We recognize the critical role of data centers in the growth of frontier technologies such as artificial intelligence and cloud computing, as well as the broader digital economy. Simultaneously, these facilities also require significant inputs of water, energy, and land, and already impact local communities' access to these essential resources. Many are facing public pushbacks and, in some cases, a real physical threat to their operations.

By examining case studies from other extractive industries, and examples of mining companies, governments, and local communities navigating water access and distribution questions, we explore the moral complexity of water as both a human right and an economic input and commodity. In addition, we call out more recent instances where such tensions have led to real consequences and impacted communities' access to safe water, and have disrupted data center operations and long-term strategy.

As a path forward, we offer a framework for public decision-making and corporate water responsibility, grounded in equity, accountability, and sustainability. This three-pillar framework carries distinct implications for local governments, water utility regulators, and companies.

Competing Values and Ethical Trade-offs

The governance of water in an era of scarcity involves genuine value conflicts rooted in one foundational premise: Water has no substitute; we are running out of it; and our personal and economic need for it is rapidly increasing.

In *The Three Ages of Water: Prehistoric Past, Imperiled Present, and a Hope for the Future*, Peter H. Gleick (2023) points out that for fossil fuels, for example, there are alternatives capable of providing comparable benefits: solar for electricity, electric motors for transportation, heat pumps for heating. For water, there are none.

The UN World Water Development Report on Valuing Water discusses the imbalance between water's price, cost, and value given conventional accounting often excludes domestic, ecological, cultural, public-health, and rights-based functions (UNESCO/ UN-Water, 2021). Water ethics scholarship similarly argues that allocation decisions embed judgments about productivity, ecology, culture, fairness, and participation, even when presented as technical management (Groenfeldt & Schmidt, 2013; Schmidt & Shrubsole, 2013).

Even the instruments most associated with a rights-based reading of water do not resolve this tension so much as encode it. The Dublin Statement (1992), often cited as an early articulation of water as a human right, in fact opens by declaring water an economic good, and only from that premise argues that essential household use should be priced, if priced at all, to protect affordability.

The United Nations General Assembly recognized safe and clean drinking water and sanitation as a human right in Resolution 64/292. The scale of the current crisis makes the ethical trade-off urgent. The World Health Organization (WHO) and the United Nations Children's Fund (UNICEF) reported in 2025 that 2.1 billion people still lack safely managed drinking water, including 106 million people who drink directly from untreated surface water.

Against this backdrop, the rapid expansion of the digital economy is manifesting itself in a physical dimension—data centers, and their resource consumption.

This is not to say that data centers are to blame for water scarcity worldwide, or that technology and artificial intelligence (AI) companies are to solve this issue alone. Rather, it is an opportunity to learn from past and recent examples on how involved parties can use an ethical lens to water governance and resource management.

In this context, it is also important to acknowledge the potential economic benefits of AI-related scaling and competition for mega-sites, grid access, and permits, and how data centers can operate as regional tech ecosystem hubs. Data centers can contribute to local economic development by generating meaningful tax revenues, construction and site operation job creation, and other indirect activity such as local services and retail sales.

Recent research by the Brookings Institution titled *Turning the Data Center Boom into Long-term, Local Prosperity* suggests that in order to reap these economic benefits, regions should treat data center development not as “isolated real estate transactions but as ecosystem-shaping moments that trade infrastructure access for commitments to advance local innovation, talent, and industry strengths.”

Economic Efficiency vs. Human Need

When water is treated primarily as an economic input, productivity gains in industry, agriculture, and digital infrastructure are achieved, but at times this occurs at the expense of affordable access for households and marginalized communities. In water-scarce settings, efficiency metrics can override basic human needs.

A data center that secures water at below-market rates to sustain operations is functioning efficiently in a narrow economic sense, but if that arrangement is achieved through opaque contracting while nearby residents face rising costs and restricted access, it raises questions about fairness, dignity, and the minimum standards of access that any society owes its members.

Corporate Certainty vs. Democratic Accountability

Private investors require predictable, long-term water access to underwrite major capital investments. Governments seeking to attract that investment may respond by offering preferential tax rates, long-term exemptions, or other benefits. These arrangements can severely limit public scrutiny and weaken democratic participation in decisions about a shared resource.

The result is a structural trade-off between attracting private capital and preserving the accountability of governments to the citizens they serve. When siting and permitting decisions for major water users are made without transparency, as has occurred repeatedly in data center development across the United States, public trust erodes and the legitimacy of both the company and the government is damaged.

Short-Term Development Gains vs. Long-Term Ecological Sustainability

Water-intensive activities can generate rapid economic returns, tax revenues, employment, and infrastructure investment. They can also degrade aquifers, rivers, and ecosystems that are the foundation of future water security for entire regions.

The ethical tension arises when policy choices shift environmental costs to future generations and to communities with limited adaptive capacity.

Resource Management: Lessons from Extractive Industries

The mining sector offers instructive precedents for how extractive industries should, and should not, engage communities over water and address the ethical complexities of water allocation. Research by Avant, Finn, and Olsen (2023) on two Peruvian mines, Cerro Verde and Yanacocha, illustrates what it means to earn, or fail to earn, a social license to operate, with findings directly relevant to data center development.

Cerro Verde: Transformational Engagement and Earned Legitimacy

When Freeport-McMoRan's Cerro Verde copper mine near Arequipa announced a 2003 expansion requiring additional water rights, it faced organized opposition led by local civil society group FREDICON over water shortages and contamination. An initial transactional response, offering discrete project benefits, failed to build trust. Cerro Verde then shifted strategy, engaging broadly with community concerns as interrelated rather than as isolated complaints, and discovered that water access and basic services, not the mine itself, were residents' central concern.

The result was a wastewater treatment plant, built in partnership with the municipal utility, that delivered clean potable water to 95 percent of Arequipa while securing the water the mine needed for expansion and rehabilitating the Río Chili. In 2015, Cerro Verde completed a \$4.6 billion expansion that tripled copper output, having earned community trust by treating the expansion as a contribution to shared development.

Yanacocha: Transactional Strategies and the Costs of Legitimacy Failure

Newmont's Yanacocha gold mine in Cajamarca followed the opposite trajectory. Community concerns over water contamination, intensified by a 2000 mercury spill in Choropampa, were met with a narrowly transactional corporate social responsibility (CSR) strategy—case-by-case grievance management rather than collective dialogue. When the International Finance Corporation (IFC) convened a multi-stakeholder process in 2001, local organizations rejected it as a “divide and conquer” tactic. Subsequent expansions triggered blockades, strikes, and violence, including the 2006 killing of farmer Isidro Llanos during protests. The proposed Conga expansion was halted entirely in 2011 after subsequent mass protests. Reactive measures like community-specific micro-reservoirs never addressed the broader water security problem, leaving lasting distrust.

Operating under the same regulatory framework with comparable resources, Cerro Verde and Yanacocha diverged on strategy and ethics: One treated community concern as inputs to a shared problem-solving process; the other as risks to be managed. For data centers navigating community opposition and water stress today, the lesson is clear: Companies that fail to genuinely earn their licenses to operate risk delays, pushback, and financial loss.

Data Centers: Implications of Water Consumption

Data centers are becoming significant industrial consumers of water, primarily for cooling for operational purposes.

According to the World Resources Institute, a mid-sized facility uses up to 300,000 gallons daily, while large hyperscale centers consume 5 million gallons per day, equivalent to a town of 30,000 to 50,000 people. By 2028, the U.S. is projected to have over 5,500 data centers. These facilities will directly consume up to 73 billion gallons of water annually, which is the equivalent of the indoor water used by about 360,000 households.

Recent corporate disclosures illustrate the scale and opacity of this consumption. [Google's 2025 sustainability report](#) showed direct water consumption of 10.9 billion gallons, a 34 percent increase from 2024, driven almost entirely by data-center cooling.

Yet direct use tells only part of the story. Indirect consumption, the water required to generate the electricity powering these facilities, varies enormously by energy source and is rarely disclosed at all. [Meta's own reporting](#) showed indirect water use in 2024 reached 19 billion gallons, more than 20 times its direct consumption, underscoring how thoroughly current accounting standards understate the true water footprint of digital infrastructure.

Some technology companies such as Nvidia and Microsoft are beginning to address this issue by promoting a closed-loop cooling design that requires no additional water once filled, a solution that both eliminates direct water draw and reduces the energy needed for cooling. However, these innovations apply only to new construction, while most existing facilities still rely on evaporative cooling systems, which are energy-efficient but water-intensive.

Water serves fundamentally different purposes for private operators and surrounding communities. For data center operators, water is a prerequisite for operational continuity—without it, servers overheat and services fail. For surrounding communities, that same water source is the foundation of health, sanitation, and agricultural stability. The tension sharpens when data centers are sited in regions where electricity is cheap, but water is already stressed, resulting in rate hikes, and water supply disruption.

Contested Claims and Absent Planning

Not all data center water conflicts look the same. Some involve competing claims over scarce resources, while others expose failures of governance before communities ever have a chance to respond.

Canelones, Uruguay: A Disputed Water Claim

Google began building a data center in Canelones in 2021. Two years later, Uruguay faced its worst drought in decades, forcing Montevideo to blend seawater into the public supply while residents struggled to access safe drinking water. When Google's planned water use became public, protests erupted under the slogan: "This is not drought, it's pillage."

Yet the conflict was more complex than it appeared. Most of Uruguay's water is consumed not by households but by export-oriented industries such as soy, rice, and pulp production, which had strained the system for years with far less scrutiny. As the newest and most visible user, Google became the focal point for broader frustrations over water allocation.

Facing public pressure, Google reduced its projected water use and adopted air cooling, and the revised project was approved. The company stayed, the community won concessions, and the underlying scarcity remained unresolved. Canelones illustrates how a rights claim can secure meaningful changes without addressing the larger economic forces driving water stress.

Newton County, Georgia: Absent Planning

The Meta data center in Newton County presents a different problem: major industrial development without baseline studies, community engagement, or adequate planning.

After construction of the \$750 million facility began in 2018, nearby residents who rely on wells reported significant water quality issues. The local development authority had conducted no pre-construction well water study, leaving no baseline data to determine whether harm occurred. Without that evidence, affected residents have little meaningful path to redress.

The consequences extend beyond individual wells. Meta's facility now uses roughly 10 percent of the county's daily water supply, while Newton County is projected to face a water deficit by 2030. Residents are expected to absorb steep water-rate increases to fund infrastructure upgrades, even as data center operators often disclose little about their water consumption.

These cases reveal opposite ends of the same governance challenge. Canelones shows what happens when a legitimate rights claim confronts a water system already shaped by larger commercial users. Newton County shows what happens when no meaningful planning or public accountability exists from the start.

Taken together, they underscore the importance of a more holistic and ethical approach to water governance.

A Values-based Framework for Ethical Data Center Management

 Our analysis calls for a coherent set of prescriptive solutions. In this section, we offer a values-based framework that emphasizes democratic process, new measurement standards, and mandatory risk and impact assessment into the governance of data center water use. Our intention is to complement existing and emerging rules and regulations rather than calling for the development of new entities or treaties.

The three pillars of this framework are mutually reinforcing. Firstly, the democratic process creates accountability that makes measurement meaningful. Secondly, measurement creates an informational foundation that makes impact assessment honest and transparent. Thirdly, impact assessment creates the forward-looking discipline that gives democratic participation genuine stakes.

Pillar One: Democratic Process and Community Participation

The core lesson of Cerro Verde is that companies do not earn legitimacy by managing community relations; they earn it by genuinely participating in the governance of shared concerns. For data centers, this requires moving from the current model of ad hoc, post-controversy community outreach toward structured, rights-based participatory processes embedded in the permitting and siting framework from the outset.

Concretely, this means that before any major water use permit is granted, a structured community engagement process must be completed, not as a formality, but as a substantive dialogue with the capacity to influence project design and conditions. This process should include direct engagement with affected households. It should also include meaningful representation of marginalized groups and renters who lack the political capital to advocate through formal channels. Moreover, it should be a non-technical presentation of proposed water use data that is open to project modification based on community input.

As a condition of permitting, communities and companies should also agree on an emergency water use hierarchy—a binding protocol that ensures drinking water, sanitation, hospitals, schools, subsistence livelihoods, and ecological minimum flows take priority over discretionary industrial use during any period of drought or acute water stress.

For example, Virginia’s Department of Environmental Quality [issues water permits with drought triggers built into the permit](#) and tied to a three-stage drought classification system. When a 2026 statewide drought hit, data centers in Henrico, Fairfax, and Loudoun counties faced the same mandatory restrictions as residential customers, with restrictions going in effect automatically once a drought stage is declared, rather than being negotiated ad hoc. Embedding this hierarchy in permit conditions before a crisis materializes is far more effective than attempting to negotiate it after one does.

The model of “local water councils,” standing advisory bodies that include affected community members in ongoing oversight of water use by large industrial facilities, offers one vehicle for institutionalizing this engagement beyond the initial permitting phase.

Democratic process also requires transparency as a precondition. The current practice, documented in multiple U.S. jurisdictions, of data center companies obscuring their water use through aliases, confidential permitting processes, or simple non-disclosure, is fundamentally incompatible with democratic governance of a public resource.

Governments should require that water use data for large industrial facilities be disclosed publicly as a condition of permitting, with no exemptions for commercial confidentiality.

Companies, for their part, should adopt the approach demonstrated by Cerro Verde: treating democratic engagement not as a risk management exercise but as an operational necessity. When a data center enters a watershed, it enters a relationship with the people who depend on that water. The quality of that relationship, whether it is transactional and defensive, or transformational and collaborative, will determine whether the company can operate sustainably.

“When a data center enters a watershed, it enters a relationship with the people who depend on that water.”

Pillar Two: New Ways to Measure & Report Water Impact

Effective governance requires transparent accounting and reporting. The current frameworks for measuring data center water use are systematically inadequate: They capture only direct operational consumption, ignore the substantially larger indirect water footprint from energy production, and rarely account for cumulative impacts when multiple large users share a stressed basin. In addition, companies are not required to report on their water usage and recycling rates either to regulatory bodies or to the public.

A values-based framework requires new measurement standards that reflect the full scope of water use and its distributional consequences. Full water footprint accounting should become standard for all major industrial water users. This means reporting not just gallons withdrawn from a specific source, but total water consumption including indirect use, incorporating the water embedded in electricity generation, construction, and supply chains. Given that indirect use may be 12 times higher than direct use, current disclosure standards are not a conservative underestimate; they are a fundamental misrepresentation of the true water cost of digital infrastructure.

Beyond volumetric accounting, measurement frameworks should incorporate equity metrics that assess distributional impacts. It must be consistent with emerging sustainability reporting approaches that require companies to disclose water-related impacts, dependencies, and management responses rather than only physical withdrawal volumes (GRI, 2018; TNFD, 2023). For instance, “water equity performance” could measure whether a company’s water use is improving or degrading access, affordability, and reliability for surrounding communities. (UN General

Assembly, 2010; OHCHR, n.d.). To avoid becoming a self-reported public relations tool, such a metric should be defined by regulators or recognized sustainability standards. It must be independently audited by accredited third parties, and subject to public disclosure. Models such as the Alliance for Water Stewardship certification system, which provides external assurance for responsible water stewardship claims, could become a standard reporting practice alongside financial performance data. Extending this scrutiny to community-level water equity would create incentives for companies to internalize the social costs of their operations rather than externalizing them onto households and municipalities.

Promising models for developing such metrics are already emerging through academic-community partnerships. [The University of Michigan Graham Sustainability Institute's](#) partnership with the regional water utility offers an instructive example, bringing together researchers, local stakeholders, and industry actors to comprehensively measure how data center siting, design, and operations affect local communities. Such partnerships demonstrate that the informational infrastructure needed for honest water accounting is achievable and provides a replicable model for other regions navigating rapid data center expansion.

Measurement frameworks should also be prospective as well as retrospective. Basin-level stress projections, updated regularly and publicly available, should inform permitting decisions. When a county is projected to face a water deficit within a decade, that information should trigger automatic heightened scrutiny for new large-scale water use permits.

Pillar Three: Mandatory Risk and Impact Assessments

The most direct lesson of the cases documented in this report is the cost of proceeding without adequate pre-construction assessment. The Joint Development Authority in Newton County did not conduct a well water study before Meta broke ground. There was no baseline against which subsequent changes in groundwater quality could be assessed, no mechanism for identifying and remedying harm, and no community information that would have allowed residents to give informed consent to the development. This is not an isolated oversight; it is the prevailing norm, and it is ethically untenable.

Environmental impact assessments (EIAs), as deployed in the Cerro Verde case, proved critical in providing a formal locus for community engagement and a mechanism through which civil society could challenge inadequate planning. The same model should be applied systematically to large-scale data center development, with specific requirements for water-focused components:

1. **Baseline groundwater surveys** for all properties within a defined radius.
2. **Cumulative impact analysis** that accounts for existing and planned water users in the basin.
3. **Equity analysis** that examines differential impacts on households, particularly those relying on private wells.
4. **Contingency planning** specifies remediation obligations if harm to groundwater quality or availability is identified.

Risk assessments should be similarly mandatory, prospective, and enforceable. A risk assessment for a proposed data center in a water-stressed basin should model worst-case scenarios for groundwater drawdown and contamination, identify the populations most vulnerable to those scenarios, and specify the conditions under which the project should be modified, conditioned, or denied. Crucially, both the risk assessment and the EIA should be completed before permitting decisions are made, not filed after the fact as documentation of decisions already taken.

Accountability requires that these assessments be binding, not advisory. Companies should be required to post financial bonds against identified risks before construction begins, creating direct economic incentives for honest assessment and adequate mitigation planning. Where harm to community water access is later documented and traceable to a company's operations, the baseline data from

the pre-construction assessment provides the evidentiary foundation for both regulatory enforcement and legal redress. Without such a foundation, affected communities are left without recourse.

The framework of mandatory pre-construction assessment, participatory engagement, and transparent disclosure is not a barrier to data center development. Cerro Verde's expansion, achieved after a genuine community engagement process that produced a shared water solution, nearly tripled the mine's copper output, and created lasting civic infrastructure that served the city of Arequipa for decades. Companies willing to do this work do not slow development; they build the foundations of durable operational legitimacy. Those that are not willing to do it should not be granted access to public water resources.

Conclusion

Water allocation is ultimately a question of values. Every permitting decision, infrastructure investment, and corporate siting choice reflects judgments about whose needs take priority, whose risks are acceptable, and what obligations the present owes to future generations. The rapid expansion of data centers has brought these questions into sharper focus, but the underlying challenge extends far beyond the technology sector.

The cases examined here demonstrate that water conflicts rarely arise from a single facility alone. They emerge when governance systems fail to account for cumulative impacts, exclude affected communities from decision-making, or treat water only as an economic input rather

than a resource essential to human dignity, public health, and ecological stability.

Hence, the challenge facing governments, regulators, utilities, and companies is not to choose between economic development and water security, but to govern both responsibly.

“The rapid expansion of data centers has brought these questions into sharper focus, but the underlying challenge extends far beyond the technology sector.”

Notes and References

Association of Water Companies (NAWC) Water Industry. <https://nawc.org/water-industry/>

Avant, Deborah, Daniel Finn, and Tricia D. Olsen. "Can CSR Strategy Mediate Conflict over Extraction? Evidence from Two Mines in Peru." *World Development* 170 (2023): 106323.

Bakker, Karen. *An Uncooperative Commodity: Privatizing Water in England and Wales*. Oxford: Oxford University Press, 2004.

Bakker, Karen. *Privatizing Water: Governance Failure and the World's Urban Water Crisis*. Ithaca, NY: Cornell University Press, 2010.

Boelens, Rutgerd, Tom Perreault, and Jeroen Vos, Eds. *Water Justice*. Cambridge: Cambridge University Press, 2018.

CBS Minnesota. "Farmington Residents Push Back Against Massive Data Center Projected to Double City's Water Use." CBS News Minnesota, February 23, 2026. <https://www.cbsnews.com/minnesota/news/farmington-hyperscale-data-center-pushback/>.

Díaz-Marín, Carlos D. "AI Data Centre Waste Heat Could Be Used for Water Purification and Carbon Capture." European Commission, Directorate-General for Environment, March 30, 2026. https://environment.ec.europa.eu/news/ai-data-centre-waste-heat-could-be-used-water-purification-and-carbon-capture-2026-03-30_en.

Food & Water Watch. *The State of Public Water in the United States*. Washington, DC: Food & Water Watch, 2016. https://www.foodandwaterwatch.org/wp-content/uploads/2021/03/report_state_of_public_water.pdf.

Gleick, Peter H. *The Three Ages of Water: Prehistoric Past, Imperiled Present, and a Hope for the Future*. London: Hachette UK, 2023.

Goetzal, Daniel, Mark Muro, and Shriya Methkuppally. "Turning the data center boom into long-term, local prosperity." Brookings Institution, February 5, 2026.

Gorey, Jon. "Data Drain: The Land and Water Impacts of the AI Boom." *Land Lines*, Lincoln Institute of Land Policy, October 17, 2025. <https://www.lincolninst.edu/publications/land-lines-magazine/articles/land-water-impacts-data-centers/>.

Groenfeldt, David, and Jeremy J. Schmidt. "Ethics and Water Governance." *Ecology and Society* 18, no. 1 (2013): art. 14. <https://www.jstor.org/stable/26269262>.

Intergovernmental Panel on Climate Change (IPCC). "Freshwater Resources." In *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by C. B. Field et al. Cambridge: Cambridge University Press, 2014. https://www.ipcc.ch/site/assets/uploads/2018/02/WGIIAR5-Chap3_FINAL.pdf.

Lei, Nuoa, Jun Lu, Arman Shehabi, and Eric Masanet. "The Water Use of Data Center Workloads: A Review and Assessment of Key Determinants." *Resources, Conservation and Recycling* 219 (2025): 108310. <https://doi.org/10.1016/j.resconrec.2025.108310>.

Madani K. (2026) Global Water Bankruptcy: Living Beyond Our Hydrological Means in the Post-Crisis Era, United Nations University Institute for Water, Environment and Health (UNU-INWEH), Richmond Hill, Ontario, Canada, doi: 10.53328/INR26KAM001.

Mims, Christopher. "AI Data Centers Use Far More Water Than Most Tech Giants Report." *The Wall Street Journal*, July 3, 2026. <https://www.wsj.com/tech/ai/ai-data-centers-water-use-901e2902>.

National Association of Water Companies (NAWC). "Water Industry Overview." Accessed 2025. <https://nawc.org/water-industry/>.

Office of the Director of National Intelligence. "Future of Water." In *Global Trends 2040: A More Contested World*. Washington, DC: ODNI, March 2021. <https://www.dni.gov/index.php/gt2040-home/gt2040-deeper-looks/future-of-water>.

Rogoway, Mike. "The Dalles Settles Public Records Lawsuit over Google's Data Centers, Will Disclose Water Use." *The Oregonian/OregonLive*, December 14, 2022. <https://www.oregonlive.com/silicon-forest/2022/12/the-dalles-settles-public-records-lawsuit-over-googles-data-centers-will-disclose-water-use.html>.

Ruggie, John. *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework*. New York: United Nations, 2011. https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf.

Schmidt, Jeremy J., and Dan Shrubsole. "Modern Water Ethics: Implications for Shared Governance." *Environmental Values* 22, no. 3 (2013): 359–79.

Solon, Olivia. "Drought-Stricken Communities Push Back against Data Centers." NBC News, June 21, 2021. <https://www.nbcnews.com/tech/internet/drought-stricken-communities-push-back-against-data-centers-n1271344>.

Sultana, Farhana, and Alex Loftus, eds. *Water Politics: Governance, Justice, and the Right to Water*. New York: Routledge, 2020.

Swyngedouw, Erik. *Social Power and the Urbanization of Water: Flows of Power*. Oxford: Oxford University Press, 2004.

Tan, Ernest. "Meta Built a Data Center Next Door. The Neighbors' Water Taps Went Dry." *The New York Times*, July 14, 2025. <https://www.nytimes.com/2025/07/14/technology/meta-data-center-water.html>.

United Nations Development Programme (UNDP). *Human Development Report 2006: Beyond Scarcity—Power, Poverty, and the Global Water Crisis*. New York: UNDP, 2006. <https://hdr.undp.org/system/files/documents/hdr2006-complete.pdf>.

United Nations Office of the High Commissioner for Human Rights (OHCHR). *The Right to Water*. Written by Catarina de Albuquerque, UN Special Rapporteur on the Human Right to Safe Drinking Water and Sanitation. Geneva: OHCHR, 2011. <https://www.ohchr.org/sites/default/files/Documents/Issues/Water/BookonRightWaterSanitation.pdf>.

United Nations Water (UN Water). *The United Nations World Water Development Report 2021: Valuing Water*. Paris: UNESCO, 2021.

WCIV / ABC News 4 (Charleston). "New Google Data Centers Promise Jobs but Raise Questions about Water Usage Impact." October 14, 2025. <https://abcnews4.com/news/local/new-google-data-centers-promise-jobs-but-raise-questions-about-water-usage-impact-dorchester-county-berkeley-county-weather-impacts-ai>.

World Resources Institute. "7 Ways Data Centers Affect US Communities." *WRI Insights*, February 16, 2026. <https://www.wri.org/insights/us-data-center-growth-impacts>.

A New National Security Architecture

An Economic Security & Strategic Investment Framework for the Western Hemisphere

JOSHUA ACOSTA

DENVER BARROWS



CREDIT: Adobe/CLUC

Introduction

The United States has been at the vanguard of democratization and economic development. During the post-Cold War era, a dominant dollar, maximum pressure campaigns, and weaker autocratic states created conditions for the United States to assert its strategic gains unchallenged. This allowed policymakers and political leaders to overlook ethical frameworks or values-based economic policies that could act as checks or guards against potential democratic decay both internally and internationally. The thinking was, if the United States had no geopolitical rival that caused an existential threat to democracy, then the need for a standard of principled economic statecraft to safeguard democratic norms is unnecessary. With the U.S. banking system at the center of the global economy, the nation had legal and functional control over the financial plumbing of the world economy,¹ not moral authority to negate democratic backsliding but governmental authority through financial control. Through issuance of presidential executive orders or congressional action Washington could leverage economic costs on foreign targets and adversaries with little to no pushback. In other words, there was no immediate need to centralize moral authority as a key tenet for economic security.

As a result of this centrality of power, **democratization became categorization and economic statecraft became a precursor to kinetic conflict.**² The categorization was simple: 1) states that are non-democratic must democratize; 2) agree to benchmarks of a U.S.-led system; or 3) risk categorization by the U.S. as either autocratic, adversarial, or fragile. If the most powerful economic authority in the world categorizes a state as adversarial or fragile, said state would become excluded from the benefits of a U.S.-led financial system. Washington utilized tools of economic statecraft as blunt instruments with little to no precedent of the spillover effects on populations within the targeted state; no unified definition of economic security, no internal or international consensus of economic statecraft, and the negation of checks and balances to help citizens of developing countries.

This created an inevitable moral void, and this lack of ethical oversight cleared the path for a degradation of analytical acumen needed to understand the realities of a shifting international system. Washington overlooked institutions in developing countries—their durability, and their structure of governance—assuming the U.S. economic security umbrella was enough to counter any threat.

¹ Daniel McDowell, “New Era of Financial Sanctions: Adapting to Anti-Dollar Policies” in *Transatlantic Economic Statecraft: Different Approaches, Shared Risk* (Atlantic Council, Washington, DC 2023), 38, <https://www.atlanticcouncil.org/wp-content/uploads/2023/09/Transatlantic-Economic-Statecraft.pdf>.

² Susan L. Woodward, *The Ideology of Failed States: Why Intervention Fails*. (Cambridge, United Kingdom; Cambridge University Press, 2017), 64-65.

This oversight allowed time and space for rising authoritarian nations to develop economic statecraft with tools of their own. China, which this report will center around, and various other non-democratic states began to leverage the linkages of a U.S.-led financial system to create economic chokepoints to illustrate that it is itself fragile and, therefore, does not possess the authority to label others as such. Beijing implemented a strategic narrative: The U.S. is extractive; it seeks to maintain its primacy, and developing states become vassal states under economic security promises. In 2013, China launched the Belt and Road Initiative with the goal of altering the international system and bringing developing states categorized by the U.S. as fragile or non-democratic under its own umbrella. A clash of systems between economic development under the banner of the U.S. or economic development under the banner of Chinese authoritarianism emerged. Beijing's argument is that states that wish to move away from the financial and military power of the U.S. can become freer under their financial and development initiatives.

This clash of systems reveals a central ethical problem in economic security and statecraft: The primacy of the U.S. and the rising power of China to challenge that primacy have caused countries with little to no resources to choose sides. Alliances become transactional and adversaries become normalized for the sake of economic gain. This report will suggest that creating a new principled economic security umbrella is necessary to counteract the growing extractive authoritarianism

of China and correct the flawed economic security agenda of Washington. To create this values-based framework, this report will illustrate how the U.S. and Chinese economic competition prolongs conflict and ignores the needs of populations at the expense of global fragmentation. Our solutions are both international and national in the form of suggesting a new set of principles, a Geneva-style convention to create a unified definition of economic statecraft, and a strategic investment fund for democratic resiliency. The international system has entered an age of economic warfare, and the shocks that follow will fundamentally reorient the global order.

“The primacy of the U.S. and the rising power of China to challenge that primacy have caused countries with little to no resources to choose sides. Alliances become transactional and adversaries become normalized for the sake of economic gain.”

Balancing Competing Values and Key Ethical Trade-offs

To understand how the United States' own actions have led to this multipolarity throughout the international financial system, it is important to trace how the nation has overlooked the very institutions it helped create. These institutions were supposed to provide a moral framework for expanded economic security throughout the international community. In a post-World War II order, the U.S. helped to build the United Nations, North Atlantic Treaty Organization (NATO), bilateral alliances in Asia, the International Monetary Fund (IMF), World Bank, and World Trade Organization (WTO). These international organizations (IOs) were founded to help incentivize countries to cooperate within the framework of multilateralism. Membership to established IOs would act as a signal that an admitted country was democratizing. This was a moral imperative post-WWII to lock in influence and balance the power scales. These institutions would act as the moral center for international development projects and peacebuilding missions. Joining an IO sets a good precedent for democratizing states or established democracies since IOs provide their members with a variety of services such as economic, governmental, and technical expertise to enhance legitimacy.³ However, multiple administrations actively sought to undermine these very institutions and preferred the power of unilateral economic coercion in the form of sanctions and tariffs. These tools are much easier to deploy to change the behavior of an adversary as compared to utilizing the multilateral process towards actual peace and security.

To ignore the multilateral process in favor of a unilateral process invokes fear of a new Monroe Doctrine in the 21st century. In 1823, President James Monroe issued the doctrine which stipulated that any coercive acts by European nations to reassert control in the Western Hemisphere would be seen as a direct act of aggression against the United States.⁴ The U.S. has since modernized this to deter China or other geopolitical competitors. China took notice and developed its geopolitical and economic motivations on the U.S. refusal to secure the multilateral process. The Belt and Road Initiative was a counter move to this oversight by the United States. Chinese President Xi Jinping prefers to redraw trade routes and cultivate new development projects where U.S. economic might and military strength cannot go. So far, 147 countries, which account for two-thirds of the world's population and 40 percent of the overall GDP, have joined. This is perhaps because no developing country wants to help the U.S. establish a preeminence. The U.S. traded the foundational pillar of democracy and multilateralism for economic coercion and therefore lost moral authority.

3 Paul Poast and Johannes Urpelainen, *Organizing Democracy: How International Organizations Assist New Democracies* (Chicago: University of Chicago Press, 2018), 25.

4 Paul Poast, "Trump's Approach to Latin America is Not Such an Outlier," *World Politics Review* (December, 2025), <https://www.worldpoliticsreview.com/trump-monroe-doctrine-venezuela-latin-america/>.

U.S. and China Competition: The Degradation of Values-Based Economic Security

There are currently around 60 to 80 active armed state or non-state conflicts in the world, and this number has steadily risen since 2024. This is precisely why the original economic security/statecraft model of multilateral institutions as protection was vital to overall peace and security: to institutionalize aiding populations in times of conflict due to shifts in regime or governance. Democratization and democratic backsliding are extremely violent processes that normally become difficult to assess analytically. However, as the U.S. became more focused on primacy given its shift away from multilateralism, China became acutely aware of the abdication and stepped in with a **crisis-lending model of economic statecraft** to protect populations from crisis and conflict.⁵ This is not to say China was or is concerned with actual international development, but it serves the purpose to showcase new leadership that can challenge the United States. The countries emerging from conflict under a developed multilateral framework would normally go seek financial assistance to democratize from the IMF, but they are now seeking it from China. China's goal is not democratization nor even stabilization; it is to prey on weak populations and fragile governments to replace their preexisting currency with China's currency which places said nation under the financial umbrella of Beijing. China uses instruments of economic statecraft to provide emergency credit to vulnerable states or populations with its primary tool being central bank currency swap

agreements. At the moment, the People's Bank of China has around 40 such deals with foreign monetary authorities;⁶ moral calamity disguised as international development. This is not to cater to the people of conflict zones, nor is it to help nations climb to sound economic policy. It is simple extraction to make these countries dependent on China's economic might.

China also implements crisis lending for a very specific reason: to improve its image with the public both domestically and internationally. It is as much a narrative strategy as it is an economic tool.⁷ The Chinese state works if it has compliance from its citizens. The Chinese Communist Party (CCP) has a history of what is called *consultative authoritarianism*, and that is greater citizen influence to increase public voice on policy issues without undermining the existing political system.⁸ Therefore, if Beijing can project a strong image abroad it assumes its citizens will comply with the state given it allows its own population to have a somewhat limited voice in the policy process. As a result, China makes inroads towards the Western Hemisphere to prove to its own people it can challenge the U.S. in its own region. Take Argentina, back in 2023: To avoid being late on its loans with the IMF, the state used China's *renminbi* to not default. This is no accident and it allows for China to emerge as a leading economic powerhouse capable of helping countries even within the Western Hemisphere.

China also is quite lenient on financing and provides multiple external financial options to help developing countries. China will allow for bailouts if said country remains part of its Belt and Road Initiative. To be clear, none of this on

5 David A. Steinberg, Selim Erdem Aytac, and Daniel McDowell, "Does crisis lending help China win friends and influence people?" in *World Development* (April 2026), 1, <https://doi.org/10.1016/j.worlddev.2026.107421>.

6 Ibid; 1.

7 Yeling Tan, David Steinberg, and Daniel McDowell, "Global economic influence and domestic regime support: evidence from China" in *Review of International Political Economy Economic Statecraft* (2025), 206-207, <https://doi.org/10.1080/09692290.2024.2402817>.

8 Rory Truex, "Consultative Authoritarianism and Its Limits" in *Comparative Political Studies* (2014), 6-7. <https://doi.org/10.1177/0010414014534196>.

any level is the CCP suddenly caring about the international community. No, this is the Chinese government attempting to replace the U.S.-led financial system. To gain financial leverage, China exploits the system of categorization that the U.S. places countries under. The United States economic statecraft formula is rather simple to predict: Categorize as democratic or non-democratic, decide whether to levy sanctions or tariffs, and then help developing economies through as long as they fall within helping American primacy. China utilizes this unethical and disastrous economic policy as an excuse for its own crisis lending. Beijing has helped Argentina solve its sovereign debt crisis, Turkey (a hybrid regime) balance their payments, and Russia avoid sanctions levied by the United States.⁹

The Vassal State Approach to Economic Statecraft

All of this to say, there is an exploitative economic security and statecraft strategy happening within the financial system and it is being waged by two of the largest economies in the world: one by the U.S. attempting to establish its democratic model and the other by China to exploit for leverage. This leads to an international system of vassal states—not alliances, not economic development, and certainly not democratization. How has the United States responded to this noticeable counter strike from China? (Washington’s methodology to abandon multilateralism in place of unilateral economic statecraft hasn’t exactly deterred Chinese aggressive economic coercion.) One does not need to look far to find the answer. The Trump administration’s 2025 National Security Strategy created core tenets around economic security, using language similar to the example of hubris that was outlined in the opening of this report.

The report states that **the U.S. foreign policy apparatus is committed to preventing “non-hemispheric competitors” from acquiring strategic assets in America or more broadly.**

In the Americas in particular the trade-off is simple: Create a coalition of aligned governments that are ideologically willing to fall in line under the economic security umbrella and trade away their own institutional resilience to countering Chinese influence.

This happened in the form of the Shield of the Americas Summit, where Argentina, Chile, Ecuador, El Salvador, Panama, and Honduras all signed an agreement. This agreement spells out essentially militarized economic security; the benchmarks were to reduce violence within their own countries (a noble goal), but with an extractive quality where these countries become benefactors within a U.S. security umbrella with the trade-off of chipping away their ability to develop democratic institutions and secure their own sovereignty as nations. There was no mention of democratic resilience through economic statecraft and no mention of helping democratic institutions become more reactive for their citizens. Instead, it is focused on an expanded executive of each Latin American country and completely disregards the democratization processes and the voices of the millions who actually live on the local level in these countries. This is simply the Monroe Doctrine in real time in the 21st century and it is deeply flawed and will not counteract China’s economic statecraft. Instead, this economic security policy removes sovereignty and refuses to address the political realities of the region. Domestic publics in Latin American countries have a deep history of U.S. military intervention so economic intervention looks no different to local populations.

⁹ Ibid; 3.

Restrictive Economic Tools and the Downslope of “Pax Americana”

Colombia, Brazil, and Mexico all did not send delegations to the Shield of Americas Summit nor do they agree with the idea of a “pax americana” economic security umbrella. These three countries represent the largest economies in Latin America. In 2023 Brazil made a deal to move away from the dollar or the financial umbrella of U.S. power. Brazilian President Lula had declared that not every country should be tied to the fate or framework of the U.S. financial system. Instead, Brazil made a deal with China to increase its use in local currencies. Yet again, Beijing learned that the weaponization of U.S. economic statecraft causes other countries to look elsewhere and China seeks to be that answer. A problem for the U.S. is that its policymakers and politicians continue to assume we live in a unipolar moment and somehow other countries should just be grateful for the help. However, not only is this a multipolar system but it is a fragmented multipolar system. The ethical frameworks that are needed to govern economic security policy tailored to the specific regions, not just universally applied, are absent in Washington expert circles. The severity of the damage this can do to democracy cannot be overlooked. The U.S. already refused to create a unified definition of economic security during the unipolar moment it had and now in a multipolar moment it requires a higher level of ethical practice that the current institutions that form the frameworks of economic statecraft simply are not equipped for. This report will use the case studies of Brazil, Cuba, and Russia in the following section to illustrate this more directly.

Maximum Pressure Campaigns: The International Financial System as Leverage

Brazil: Brazil did not attend the Shield for Americas Summit for one very important reason which speaks to the moral oversight at play here with U.S. economic security; that is *democratic institutional stability*. Brazil held its first ever democratic elections in 1989 and created its first democratic constitution in 1988 with the sole goal of *curbing inflation, opening free trade, and developing the tools necessary to help the public develop their own economic opportunities*.¹⁰ It centered its own democracy around economic development which differs from the U.S. economic security framework of falling in line. It is no wonder President Lula decided in 2023 to make a deal with China to promote local currencies; it is literally how Brazil practices democracy. As mentioned earlier, U.S. economic policy is so focused on maintaining its primacy and countering China that it fails to even develop the ethical or analytical muscles needed to understand the core tenants of another country.

Cuba: The current administration’s end goal is to exert maximum pressure on the Cuban government with the purpose of regime change. This is to prove to adversaries such as China that the U.S. possesses hemispheric influence even over countries with longstanding adversarial tension. Cuba itself is undergoing an energy crisis which causes its own people to suffer the consequences. Instead of finding a diplomatic solution, Washington has escalated sanctions to further erode its energy infrastructure which damages pre-existing

10 Amory Gethin, et. al, *Political Cleavages and Social Inequalities: A Study of Fifty Democracies, 1948-2020* (Boston, Massachusetts, Harvard University Press, 2021), 438.

institutions and hollows out any chance of economic growth for Cuban citizens. However, like Brazil there is a lack of understanding of the institutions within the nation itself. The Cuban government was created with the goal of the ruling party controlling all aspects of Cuban life including its economy and government. Similar to other revolutionary governments, Cuba's regime radically transformed its institutions at home, promoted anti-democratic revolutions abroad, closed ranks to defend the regime, and built loyal security forces to wipe out centers of societal power.¹¹ The state-society power asymmetry transfers economic pain to its own populations. The current U.S. economic statecraft strategy overlooks this history. Internal collapse, which is needed for actual regime change, is unlikely to occur at all if U.S. military and economic statecraft strategy ignores these realities. Instead, if the United States continues its maximum pressure campaign it will harm Cuban citizens which the regime is content to allow as long as it maintains control over the political system. This is another example of U.S. economic statecraft being weaponized without a coherent plan beyond countering China and ensuring dominance.

Russia: Russia's experience with sanctions has been layered. In 2014, a year after China announced their Belt and Road Initiative, Russia was hit with financial sanctions as a result of its illegal annexing of Crimea. Then again in 2018, the Trump administration targeted key Russian oligarchs, officials, and export sectors. However, Russia, like China, adapted. During this layered process of economic statecraft in 2018-19, Russia decided to adjust its foreign exchange reserves and cut its dollar holdings from 44 percent to 23 percent. The Russian government shifted towards the Chinese *yuan* as a currency of choice instead. What this shows is even against an adversary that deserves maximum pressure, ethics still must be applied or spillover costs incur. Russia and China, two economic powers and anti-democratic forces, became more aligned as a result of this and the U.S. was powerless to stop it. If the U.S. took an ethical realist approach to economic statecraft it would have the framework and knowledge to understand that anti-democratic governments prefer economic hardship on their own populations to further entrench themselves within broader authoritarian systems.

¹¹ Steven Levitsky, and Lucan Way. *Revolution and Dictatorship: The Violent Origins of Durable Authoritarianism*. Princeton (New Jersey: Princeton University Press, 2022), 203.

The Necessity of a Strategic Investment Framework

The United States and China cannot compete geopolitically and yet advance economic security gains without something in the international financial system breaking both governmentally and globally. The United States would be smart to pivot to regional empowerment. What this means is that the global economic security apparatus needs to build more issue-specific coalitions. Maintaining dominance is not only morally incoherent with governance, it is also logistically impossible in the long run. Populations of developing economies—take for instance, Brazil—need to feel they are stakeholders in the financial system not just vassals in it. As a result, we propose strategic investment strategies in the form of an international investment multilateral response that involves populations and their governments most affected by economic warfare, then domestically a sovereign wealth fund (SWF) that allows for the protection of the state itself. The U.S. needs to build its institutions to allow for a two-front ethical framework that accounts for a shift in political and power dynamics. We first will map out the need to protect the state itself, then explain the necessity of a sovereign wealth fund, and then suggest an ethical framework for a new international economic security regime.

Disillusionment Amongst Citizens: Diagnosing the Moral Lapse of the U.S. Government Amidst Chinese Competition

In its quest to create an economic statecraft framework as a coercive force on the world stage, American economic policy has forgotten its core obligation: to provide a secure foundation on which every American has a pathway toward economic success. This is not a policy preference, but the precise premise the United States was built on. The American tradition once treated economic security as inseparable from legitimate authority. But that contract has been broken, and the consequence is a deep and measurable distrust of American institutions. The clearest evidence is not only in the economic opportunity Americans have lost, but in the years of life they have lost with it. Mortality data showed this before polling did; when working-class Americans, concentrated in the manufacturing communities hollowed out by decades of trade and offshoring decisions, died in a way that researchers came to call “deaths of despair.”¹² Washington treated these communities as a footnote, immaterial to the story of what America must look like, and to the trade policy being made. It decided whose flourishing counted, and that neglect is what left the country exposed to China. The same neglect shaped how the United States approached China. It showed that a state that will not secure its own citizens’ prospects will not notice, until it is too late, that it has also stopped securing its own capability—and China’s economic strategy was built to

12 Anne Case and Angus Deaton. *Deaths of Despair and the Future of Capitalism* (Princeton, New Jersey, Princeton University Press, 2020).

exploit exactly that inattention. Beijing used economic openness as an input to state design. In exchange, Washington accepted cheaper goods on American shelves and market access for American firms and treated that convenience as sufficient compensation for the capability it was surrendering.

Chinese state capital moved deliberately into the sectors, like batteries, critical minerals, and advanced manufacturing, that the United States had left to the market. And they paired it with the extraction of intellectual property built by Americans through forced transfer (i.e. joint ventures) and theft.¹³ The result is a form of dependency the public did not choose and was never asked to formally support. Policymakers made this happen through a series of individually defensible choices—in many cases at the behest of business leaders—without asking whether the citizens absorbing the risk had agreed to the bet. Households absorbed pandemic-era shortages and price shocks while workers found themselves competing against a state rather than a company. A decade after the China shock began, the workers and communities it hit hardest still haven't recovered.¹⁴ The proof sits squarely at the negotiating table: The United States seeks access to critical minerals and offers soybeans in return. The disparity in technical complexity says everything about who holds leverage, and what decades of neglected obligation cost.

That same leverage problem is not limited to any single negotiation, as shown by how the nations respectively treat chokepoints. A chokepoint is where an adversary's decision, rather than the market's, determines what an

economy can and cannot do. China built its chokepoints deliberately, directing state capital into critical minerals processing, maritime logistics, and semiconductor supply chains precisely because it understood that whoever controls the bottleneck controls the leverage.¹⁵ The United States left the same sectors to the market without a plan to cover the exposure. Matching China's investments sector for sector will not solve this. What the investment is for, is what makes the difference. China's state capital serves the state first, extracting leverage abroad and consolidating control at home, with little obligation to the workers or communities whose labor built the advantage. An American strategic investment framework, if it is to counteract that model rather than imitate it, has to be built on the opposite premise: that the state's power to invest exists because citizens consented to it, and that consent is honored only if the investment returns something to the people who bear its risk.

The public's message is clear that something must change. Americans are not seeking isolationism. Majorities across party lines still want the United States engaged in the world, and a majority now supports the government taking ownership stakes in firms critical to national security.¹⁶ But that support carries a condition. Engagement abroad must deliver something at home. That is the trade-off the last several decades avoided. It sought the efficiency of open engagement against the obligation to secure the economic opportunity of the people whose consent makes that engagement legitimate. When people watch their neighbors struggle and reckon with the

13 Patrick McGee, *Apple in China: The Capture of the World's Greatest Company* (New York: Scribner, 2025).

14 David H. Autor, David Dorn, and Gordon H. Hanson, "The China Shock: Learning from Labor-Market Adjustment to Large Changes in Trade," *Annual Review of Economics* 8 (2016), <https://doi.org/10.1146/annurev-economics-080315-015041>.

15 Justin G. Muzinich, Gina M. Raimondo, James D. Taiclet, and Jonathan E. Hillman, U.S. Economic Security: Winning the Race for Tomorrow's Technologies, Independent Task Force Report No. 83 (New York, New York: Council on Foreign Relations, 2025). <https://www.cfr.org/event/cfr-task-force-report-us-economic-security-winning-race-tomorrows-technologies>.

16 *Reagan National Defense Survey* (Washington, DC: Ronald Reagan Presidential Foundation and Institute, 2025), 6, <https://www.reaganfoundation.org/reagan-institute/centers/peace-through-strength/survey/2025-reagan-national-defense-survey>.

real prospect that the next generation lives worse than this one, there is no tolerance left for a foreign policy that does not deliver for them. China's own strategy, whatever its costs abroad, is never ambiguous about who it was built to serve at home. Its own economic security policy raised its population's standard of living even as it extracted advantage from others. The United States government cannot claim the same. If economic security is the purpose citizens grant the government the power to pursue, then the instruments of that power must be built to serve that purpose. Regaining that trust, and with it the legitimacy democracy depends on, requires building those instruments deliberately. What follows is an argument for what that framework could look like.

Sovereign Wealth Fund: Institutional Resiliency through Ethical Governance

Autocracies such as China do not mind if the consequences of economic statecraft shift to the burden of its own people. The United States should be different and protect its own citizens from spillover damages of economic conflict. A sovereign wealth fund can protect and elevate its citizens. A sovereign wealth fund is not, on its own, an economic security strategy. It is one instrument by which a government can show that a state's legitimate power depends on securing the economic opportunity of the citizens who grant it this authority. Economic security is itself an ethical obligation the state owes in exchange for the consent of the governed. Resiliency, in this context, is not a matter of capital or scale. It

is a matter of whether the fund's governance can be trusted to outlast any administration that created it. Seen from that perspective, the Trump administration¹⁷ and Bernie Sanders¹⁸ are right—both have advocated for an American sovereign wealth fund but for different reasons. And what can enhance both propositions is an ethical governance built into the fund's design rather than left to be worked out later.

This is not industrial policy under a new name. It instead addresses market failures private capital cannot solve alone, where investment horizons are too long and strategic returns cannot be internalized.¹⁹ And it meets adversary state capital on the field where that competition is actually happening, without a sanction or a threat. The democratic mandate for this is not hypothetical. Most Americans already support government ownership stakes in firms critical to national security.²⁰ That support is consent, and consent is what a fund's governance must honor.

Statutory creation is usually framed as a durability argument, and it is one, but durability here is inseparable from ethics. Congress has built durable institutions before, and their architecture offers a starting point for a sovereign wealth fund. A U.S. fund needs independence, transparent accountability, and a mandate narrow enough to enforce and broad enough to matter. How it is capitalized is itself a moral choice, not a technical one. Funds capitalized by transferring state assets outside legislative oversight tend to escape the checks meant to constrain them, operating as a spending vehicle answerable to whoever controls the executive rather than to elected

17 Executive Order, "A Plan for Establishing a United States Sovereign Wealth Fund," February 3, 2025. <https://www.whitehouse.gov/presidential-actions/2025/02/a-plan-for-establishing-a-united-states-sovereign-wealth-fund/>.

18 Bernie Sanders, "A.I. Is a Public Resource. You Should Own Half of It," *The New York Times*, June 1, 2026, <https://www.nytimes.com/2026/06/01/opinion/artificial-intelligence-bernie-sanders.html>.

19 Jodi Vittori and Lakshmi Kumar, eds. *Sovereign Wealth Funds: Corruption and Other Governance Risks*. (Washington, DC: Carnegie Endowment for International Peace, 2024), 25, [https://assets.carnegieendowment.org/static/files/Vittori-Kumar-SWF%20\(2\)-1.pdf](https://assets.carnegieendowment.org/static/files/Vittori-Kumar-SWF%20(2)-1.pdf).

20 *Reagan National Defense Survey* (Washington, DC: Ronald Reagan Presidential Foundation and Institute, 2025), 6, <https://www.reaganfoundation.org/reagan-institute/centers/peace-through-strength/survey/2025-reagan-national-defense-survey>.

representatives. Funds capitalized through the sale of citizenship or residency create a structural conflict: the incentive to vet buyers competes directly with the incentive to keep the money flowing.²¹ Proposals to fund an American SWF through tariff revenue or visa sales carry the same flaw. Any stream that bypasses Congress's power of the purse builds a parallel budget by another name, and a fund meant to earn public trust cannot begin by circumventing the public's own representatives.

The Federal Reserve's Board of Governors offers the strongest architectural precedent available because it was built to survive politics rather than serve it. Presidentially appointed and Senate-confirmed, with fixed terms that outlast any single administration, its structure treats independence as a governance principle. Oversight should rest with existing financial regulators and with mandatory reporting to Congress that creates accountability without turning the fund into an appropriations vehicle, the fund's most acute vulnerability. This is not corruption in the ordinary sense, but capture, the slow conversion of a national security instrument into a mechanism for rewarding home districts and allies. History is not short on examples of sovereign funds diverted through shell companies and complicit intermediaries into the pockets of the people meant to be overseeing them.²² Anti-corruption provisions with real statutory teeth are what separates a sovereign wealth fund from a slush fund. Underneath all of this sits the question: Can the American people trust what the fund invests in, and do they have real recourse if it fails?

Resiliency without trust is just durability, and durability alone has never been enough to justify power. The international governance frameworks that exist for sovereign funds are voluntary and carry no enforcement mechanism, and the record of compliance with even self-imposed standards is thin across the funds that dominate global assets. Recourse cannot be instant. Investment cycles run long, and the returns this fund is built for may take a decade to show up. What can be built instead is a statutory review mechanism, a mandatory reauthorization measured against the fund's stated objectives that keeps citizens inside the evaluation process without forcing the fund to chase the wrong time horizon. Sovereign wealth funds reflect the quality of governance of the states that sponsor them. A fund built for democratic resilience earns that description not through its balance sheet, but through the ethical governance it cannot escape.

21 Jodi Vittori and Lakshmi Kumar, eds., *Sovereign Wealth Funds: Corruption and Other Governance Risks* (Washington, DC: Carnegie Endowment for International Peace, 2024), [https://assets.carnegieendowment.org/static/files/Vittori-Kumar-SWF%20\(2\)-1.pdf](https://assets.carnegieendowment.org/static/files/Vittori-Kumar-SWF%20(2)-1.pdf).

22 Ibid; 17.

Ethical Framework: Principled Economic Security and a New Values-Based Definition

Now that protection of the state has been established in the form of a SWF and given that China has weaponized its economic leverage, this report calls for a new ethical framework that centers around a principled approach to economic security and statecraft. The attempts to coax and exploit the interconnectedness of the world economy are too important to leave to the competition of governments that wish to assert control over the economic levers of power. These principles will create a unified definition of economic security on an international level, guide why and how economic statecraft should be deployed, and sustain the credibility of democracy in order to update the rules of the international economic order. Then we will shift towards mapping out how to utilize strategic investment as a force multiplier to ensure geopolitical competition does not harm economic security or international development.

Pillars and Principles of Economic Statecraft for our Allies and Adversaries

First Principle: Anticipate the adverse trigger events that occur as a result of economic statecraft. Ethics should be actionable not just theoretical in this instance. China acts as a lender of last resort, and therefore, steps in when a developing country or state feels unfairly managed by the United States. The objective of Washington should be coalition building and coordinating the use of economic statecraft. Multilateralism and economic statecraft should be intertwined not decoupled. Coordinating efforts with partners and key stakeholders would allow for the institutional resiliency that is necessary to offset needing assistance from autocratic regimes.

Second Principle: Economic statecraft for the sake of maximalist goals such as regime change hardly ever works and should be removed from the calculus of decision-making. Instead, economic statecraft tools should be used to change the malign behavior of an adversary to achieve a diplomatic solution. Sending political messages through economic tools such as sanctions or tariffs often invokes a response to lean heavily towards siding with Beijing over Washington. A higher bar of scrutiny is needed to minimize the hawkish nature of maximalist goals.

Third Principle: Emerging ethical issues will arise as economic warfare continues, in the sense that governments inherently serve their own security objectives first and are concerned with resiliency second. The U.S. should seek to revitalize the very institutions it built to handle emerging ethical issues of economic development such as when and how to build democratic systems in a post-conflict setting. The IMF for example should be at the forefront of crisis financing not allowing China to act when it prefers. This would require buy-in from our allies and can be achieved if Washington coordinates economic policy instead of unilaterally deploying tools of statecraft.

Fourth Principle: A unified definition of economic security must be established throughout the international system. A Geneva-style convention should be called for that allows for a rigorous dialogue to frame economic security issues from both new democracies and established democracies. Economic security by definition should be the protection of the international financial system from autocratic forces that seek to exploit the interconnectedness as a result of globalization.

“The attempts to coax and exploit the interconnectedness of the world economy are too important to leave to the competition of governments that wish to assert control over the economic levers of power.”

Conclusion: Democratization, Economic Chokepoints, and the Future of National Security

The United States has been the leading force for democratization, innovation, growth, and international development. However, a national security strategy that has failed to factor in changing coalitions, evolving economic conditions, and a nuanced understanding of multipolarity has allowed for an increase in adversarial competition. This report has called for a principled economic security strategy that seeks to understand unanticipated risks and costs to both the American public and the citizens of the countries affected by prolonged economic warfare. Recognizing economic vulnerabilities can lead to more strategic and targeted investments domestically and internationally. Non-democratic and adversarial governments are developing templates to take advantage of economic linkages within the international system that could harm the United States. This is precisely why investment in key sectors such as critical minerals, maritime logistics, semiconductors, and pharmaceutical supply chains matters ethically because its efforts will benefit the same communities hollowed out by decades of trade and offshoring decisions.

Rebuilding capability is not a side benefit. A framework for the United States that allows it to counteract future chokepoints while navigating the spillover costs to a community suffering because of restrictive economic tools should be factored into the calculus of decision-making. This is what counteracting Chinese economic policy actually requires; not just matching its investments sector for sector, but proving that a democracy can build strategic capability without treating its own citizens as expendable to the effort. This is the wager China’s model has always made and the wager this report argues the United States cannot afford to match. A sovereign wealth fund gives this framework its instrument. Ethical governance gives it its legitimacy. Investment made on these terms is not merely sound policy. It is the state beginning to make good on the contract it broke. The principles of economic security and statecraft must be centralized to ensure that the U.S. can lead during a prolonged period of economic warfare. Given the current insistence on adversaries such as China to weaponize economic chokepoints and develop more sophisticated frameworks of governance, a clash of systems is inevitable. This inevitability requires tools of economic statecraft and strategic investments that ensure democracy is resilient.

Notes and References

Autor, David H., David Dorn, and Gordon H. Hanson. "The China Shock: Learning from Labor-Market Adjustment to Large Changes in Trade." *Annual Review of Economics* 8 (2016): 205-240. <https://doi.org/10.1146/annurev-economics-080315-015041>.

Case, Anne, and Angus Deaton. *Deaths of Despair and the Future of Capitalism*. Princeton, New Jersey: Princeton University Press, 2020.

Donovan, Kimberly, Mai Nikoladze, Nicole Goldin, Mrugnak Bhusari, et al. *Transatlantic Economic Statecraft: Different Approaches, Shared Risk*, Washington, DC: Atlantic Council, 2021. <https://www.atlanticcouncil.org/wp-content/uploads/2023/09/Transatlantic-Economic-Statecraft.pdf>.

Executive Order, "A Plan for Establishing a United States Sovereign Wealth Fund," February 3, 2025. <https://www.whitehouse.gov/presidential-actions/2025/02/a-plan-for-establishing-a-united-states-sovereign-wealth-fund/>.

Feldstein, Steve, and Jodi Vittori. "Trump's Sovereign Wealth Fund Brings High Stakes and Serious Risks." Washington, DC: Carnegie Endowment for International Peace, 2025. <https://carnegieendowment.org/research/2025/04/trumps-sovereign-wealth-fund-brings-high-stakes-and-serious-risks>.

Gethin, Amory, Clara Martinez-Toledano, Thomas Piketty, *Political Cleavages and Social Inequalities: A Study of Fifty Democracies, 1948-2020*, Boston, Massachusetts, Harvard University Press, 2021.

Levitsky, Steven, and Lucan Way. *Revolution and Dictatorship: The Violent Origins of Durable Authoritarianism*. Princeton, New Jersey: Princeton University Press, 2022.

McGee, Patrick. *Apple in China: The Capture of the World's Greatest Company*. New York: Scribner, 2025.

Muzinich, Justin G., Gina M. Raimondo, James D. Taiclet, and Jonathan E. Hillman. U.S. Economic Security: Winning the Race for Tomorrow's Technologies. Independent Task Force Report No. 83. New York, New York: Council on Foreign Relations, 2025. <https://www.cfr.org/event/cfr-task-force-report-us-economic-security-winning-race-tomorrows-technologies>.

Poast, Paul. "Trump's Approach to Latin America is Not Such an Outlier" *World Politics Review*. December, 2025. <https://www.worldpoliticsreview.com/trump-donroe-doctrine-venezuela-latin-america/>.

Poast, Paul and Johannes Urpelainen, *Organizing Democracy: How International Organizations Assist New Democracies*. Chicago: University of Chicago Press, 2018.

Reagan National Defense Survey. Washington, DC: Ronald Reagan Presidential Foundation and Institute, 2025. <https://www.reaganfoundation.org/reagan-institute/centers/peace-through-strength/survey/2025-reagan-national-defense-survey>.

Sanders, Bernie. "A.I. is a Public Resource. You Should Own Half of It." *The New York Times*, 2026. <https://www.nytimes.com/2026/06/01/opinion/artificial-intelligence-bernie-sanders.html>.

Steinberg, David, Selim Erdem Aytac and Daniel McDowell. "Does Crisis lending help China win friends and influence people?" *World Development* (2026): 1-11. <https://doi.org/10.1016/j.worlddev.2026.107421>.

Tan, Yeling, David Steinberg, and Daniel McDowell, "Global economic influence and domestic regime support: evidence from China." *Review of International Political Economy* (2025), 202-225. <https://doi.org/10.1080/09692290.2024.2402817>.

Truex, Rory. "Consultative Authoritarianism and Its Limits" *Comparative Political Studies* (2014): 1-33. <https://doi.org/10.1177/0010414014534196>.

Vittori, Jodi, and Lakshmi Kumar, eds. *Sovereign Wealth Funds: Corruption and Other Governance Risks*. Washington, DC: Carnegie Endowment for International Peace, 2024. [https://assets.carnegieendowment.org/static/files/Vittori-Kumar-SWF%20\(2\)-1.pdf](https://assets.carnegieendowment.org/static/files/Vittori-Kumar-SWF%20(2)-1.pdf).

Woodward, Susan L. *The Ideology of Failed States: Why Intervention Fails*. Cambridge, United Kingdom; Cambridge University Press, 2017.

Redefining the Frontline

AI, Healthcare, and Non-State Power

REBECCA DISTLER

KATHY KIM



CREDIT: Adobe/Duskcraft

Introduction

State of AI in Healthcare

Artificial intelligence (AI) has rapidly shifted from specialized technology to general-purpose infrastructure and is becoming foundational to how people access information and conduct their day-to-day work. McKinsey's 2025 State of AI survey found that 88 percent of respondents reported regular AI use in at least one business function, up from 78 percent the prior year.¹ Pew Research Center found that 34 percent of U.S. adults had used ChatGPT as of early 2025, twice what was reported in 2023.² Among adults under 30, that figure was 58 percent.

Healthcare is one of the most visible and highly sensitive areas where this transition is unfolding. Around the world, health systems face workforce shortages, physician burnout, and unequal access to care, and are seeking cost-effective technologies to help solve these challenges. In a world where healthcare has traditionally been slow to adopt and scale new technology, AI appears to break this pattern. In 2026, the American Medical Association reported that over 80 percent of physician respondents currently use AI in a professional context, double what was reported in 2023.³

While AI has the potential to improve the efficiency, quality, and accessibility of healthcare, it is not without risks. Ethical considerations related to the development and deployment of AI in healthcare have long been documented, including concerns around bias and unrepresentative training data, privacy, data governance, accountability, liability, and the importance of human oversight.⁴

Existing ethical frameworks for AI in healthcare, as well as broader ethical AI frameworks, provide important organizing principles to guide decision-making. Yet many were developed before the widespread use of large language models (LLMs), and before the emergence of the advanced reasoning, multimodal, and agentic systems that are defining the next era of AI.

1 Singla, Alex, Alexander Sukharevsky, Bryce Hall, Lareina Yee, Michael Chui, and Tara Balakrishnan. "The State of AI in 2025: Agents, Innovation, and Transformation." McKinsey & Company, November 5, 2025. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>

2 Sidoti, Olivia, and Colleen McClain. "34% of U.S. Adults Have Used ChatGPT, About Double the Share in 2023." Pew Research Center, June 25, 2025. <https://www.pewresearch.org/short-reads/2025/06/25/34-of-us-adults-have-used-chatgpt-about-double-the-share-in-2023/>

3 American Medical Association. 2026 Physician Survey on Augmented Intelligence. Chicago: American Medical Association, Center for Digital Health and AI, March 2026. https://www.ama-assn.org/sites/ama-assn.org/files/2026-03/physician-ai-sentiment-report_0.pdf

4 Morley, Jessica, Caio C. V. Machado, Christopher Burr, Josh Cows, Indra Joshi, Mariarosaria Taddeo, and Luciano Floridi. "The Ethics of AI in Health Care: A Mapping Review." *Social Science & Medicine* 260 (2020): 113172. <https://doi.org/10.1016/j.socscimed.2020.113172>

That next phase may be arriving sooner than anticipated. AI models have already passed medical licensing exams and performed well on structured clinical assessments.^{5,6} More recent research suggests that AI systems are beginning to meet or exceed physician performance on more complex tasks. A 2026 Science study, for example, found that an LLM outperformed physician baselines on real-world emergency department case studies, a fundamentally harder challenge.⁷ These findings should not be interpreted as evidence that AI can or should replace clinicians, but they are signals of how capabilities are rapidly evolving—and of the new ethical considerations those capabilities should raise, even while deployment decisions must still depend on validation, real-world outcomes, and continuous monitoring.

Understanding this becomes more urgent when considering that non-state actors primarily lead the development of these technologies. AI labs, health technology companies, vendors, and others are increasingly shaping which problems AI is designed to solve, how this technology is integrated into health systems, and who is able to access them.

Against this backdrop, this paper explores a set of critical questions. First, as AI systems begin to reliably outperform humans in healthcare-related tasks, how do the ethical considerations shift and how should current frameworks guide this next phase? Second, assuming these capabilities and the evidence of their impact continue to advance, what new ethical obligations might arise in relation to the non-state actors that develop these systems, and in the settings where the potential health impact is the greatest?

“As AI systems begin to reliably outperform humans in healthcare-related tasks, how do the ethical considerations shift and how should current frameworks guide this next phase?”

5 Tiffany H. Kung et al., “Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models,” PLOS Digital Health 2, no. 2 (2023): e0000198, <https://doi.org/10.1371/journal.pdig.0000198>

6 Karan Singhal et al., “Large Language Models Encode Clinical Knowledge,” Nature 620 (2023): 172–180, <https://doi.org/10.1038/s41586-023-06291-2>

7 Peter G. Brodeur et al., “Performance of a Large Language Model on the Reasoning Tasks of a Physician,” Science 392, no. 6797 (2026): 524–527, <https://doi.org/10.1126/science.adz4433>

Ethics and Power in the Age of AI in Healthcare

AI Ethics Frameworks for Health and Medicine

While AI has been used in medical practice since the 1970s,⁸ the attempt to govern it through a layered ecosystem of ethical, technical, and legal frameworks is a more recent endeavor. In 2021, the World Health Organization (WHO) published the first major global framework focused specifically on the ethics and governance of AI for health. Since then, major associations such as the American Medical Association (AMA) and National Academy of Medicine (NAM) have followed with their own ethics frameworks and principles.

Table 1. Summary of AI Ethics Frameworks for Health and Medicine

Framework	Publishing Institution	Overview
Ethics and Governance of Artificial Intelligence for Health (2021) ⁹	World Health Organization (WHO)	The first major global health-specific framework on the ethics and governance of artificial intelligence; identifies ethical risks in health applications and sets out six principles for responsible use: human autonomy, safety, transparency, accountability, equity, and sustainability.
Trustworthy Augmented Intelligence in Health Care (2024) ¹⁰	American Medical Association (AMA)	Physician-facing AI principles and policy framework emphasizing ethical, equitable, responsible, accurate, and transparent AI.
Statement on Artificial and Augmented Intelligence in Medical Care (2025) ¹¹	World Medical Association (WMA)	Global physician ethics statement; emphasizes patient welfare, physician responsibility, transparency, safety, privacy, equity, professional judgment, and responsible use of AI in medical care.
An Artificial Intelligence Code of Conduct for Health and Medicine (2025) ¹²	National Academy of Medicine (NAM)	A health and medicine-specific code of conduct intended to align stakeholders around responsible development and use of AI.

8 Casimir A. Kulikowski, “Beginnings of Artificial Intelligence in Medicine,” *Yearbook of Medical Informatics* 28, no. 1 (2019): 263–270.

9 World Health Organization, *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance* (Geneva: World Health Organization, 2021), <https://www.who.int/publications/i/item/9789240029200>

10 American Medical Association, *Augmented Intelligence Development, Deployment, and Use in Health Care* (Chicago: American Medical Association, November 2024), <https://www.ama-assn.org/system/files/ama-ai-principles.pdf>

11 World Medical Association, “WMA Statement on Artificial and Augmented Intelligence in Medical Care,” adopted by the 76th WMA General Assembly, Porto, Portugal, October 2025, <https://www.wma.net/policies-post/wma-statement-on-artificial-and-augmented-intelligence-in-medical-care/>

12 National Academy of Medicine, *An Artificial Intelligence Code of Conduct for Health and Medicine: Essential Guidance for Aligned Action*, ed. Laura Adams, Elaine Fontaine, Michael Matheny, and Sunita Krishnan, NAM Special Publication (Washington, DC: National Academies Press, 2025), <https://doi.org/10.17226/29087>

These existing frameworks converge around a core set of principles. For example, all four frameworks emphasize safety, often grouping this principle with others such as “accuracy and efficacy” (WHO) and “high-quality” (AMA) that reflect the importance of ensuring AI models work as intended in healthcare contexts. Notably, NAM also links safety to preventing “potential harm.” In prioritizing safety, they also call for post-deployment and performance monitoring and ongoing assessment in real-world use.

All four frameworks also recognize AI’s potential to both reduce and potentially deepen health inequities. NAM, for example, explicitly calls for “support to low-resourced organizations and communities to ensure equitable access to the benefits of AI,” and WMA states that “new and beneficial AI healthcare tools must be developed and deployed equitably, with the goal of being accessible worldwide.”

While these frameworks provide a solid foundation for governing evolving AI capabilities, they also primarily conceive of AI as a tool that supplements human decision-making. Furthermore, they do not contemplate the potential of future autonomous systems that could take on more end-to-end clinical tasks.

As drafted, all four frameworks rely heavily on **human oversight** as a primary safeguard for AI use in healthcare. In fact, the WHO names “protect autonomy” as its first principle precisely because “people must remain in control of health-care systems and medical decisions.”¹³ AMA intentionally uses “augmented intelligence” to reflect AI’s “assistive role, emphasizing that its design enhances human intelligence rather than replaces it.” WMA goes furthest with its emphasis on “physician-in-the-loop,” stating that a licensed physician should “review and retain final authority over all AI

outputs before they shape clinical care,” to help mitigate risk and ensure patient safety.

The future of healthcare must remain human-centered. Yet, human-centered care will not always require the same level of human review of every AI-enabled action. As systems become more capable, ethical frameworks will need to calibrate oversight to level of risk—for example, requiring direct review for high-stakes decisions while allowing systems-level supervision and ongoing monitoring where appropriate.

Physicians, nurses, and other health practitioners contribute critically important cultural and contextual understanding, empathy, and relational care that AI systems should, no matter how capable, always be designed to amplify. As AI systems evolve, the question is whether every use of AI will always require the same form or intensity of human oversight, and how to evaluate and make decisions about when to move that proverbial goal post.

Today, continuous human oversight throughout the AI lifecycle is critical for ensuring that models perform as intended and that we continue to steer the development of the technology appropriately. But as increasingly powerful models become more widely available, we will need to ensure that the level of human review is proportional to the use case or decisions resulting from its use. Otherwise, AI use in critical environments could be impacted by process and overstandardization, leading to slower access to care, false assurances, administrative burden, or the reintroduction of human error into processes where the AI system was already proven to perform more reliably.

13 World Health Organization, Ethics and Governance of Artificial Intelligence for Health: WHO Guidance (Geneva: World Health Organization, 2021), 27, <https://www.who.int/publications/i/item/9789240029200>

This is not necessarily the state of technology today, but it is a potential future state we believe our ethical frameworks need to be prepared to address. Rather than treating human oversight as a one-size-fits-all or standardized safeguard, we need a more nuanced approach to understanding oversight of a given system, one that takes into account the system's demonstrated performance, baseline alternatives, vulnerability of affected populations, and the ability of patients and clinicians to question or override its use.

In this context, **accountability** also merits renewed consideration. All four frameworks center accountability as a core ethical principle. WHO, for example, frames accountability at a systems-level, stating that AI governance should hold “all stakeholders—in the public and private sector—accountable and responsive” to health workers, communities, and affected individuals. NAM does as well, defining accountable AI as the need to “document actions, benefits, and safeguards that ensure responsibility for AI’s effects,” effectively placing accountability not just on the clinician using the tool but on the broader system that develops, purchases, and deploys it.

On the other end of the spectrum, AMA focuses more on protecting physicians, working to ensure that they are not held responsible for tools or administrative systems they do not control. WMA attempts to do both:

emphasizing that the “final responsibility and accountability” for diagnosis and therapy must lie with the physician, while also calling for “clearly distributed accountability” and “shared and proportional” liability.

While these frameworks emphasize that AI in healthcare should remain accountable, equitable, and focused on human well-being, their treatment of oversight and accountability also raises unresolved questions. As future AI systems evolve from assisting clinicians to performing defined and complex tasks, the ethical challenge will be to clarify when human control is necessary versus when it is merely symbolic and how responsibility should be allocated across the actors who design and deploy these systems.

Many of the most consequential decisions about AI-enabled care will not necessarily be made by physicians or patients. They will be made upstream by technology companies, model developers, and other non-state actors—including philanthropies, payers, and healthcare institutions—that shape which tools get built, where they are deployed, and who benefits from their use.

The question, then, is what ethical obligations fall on the non-state actors that develop and control the infrastructure of powerful AI in healthcare.

Non-State Power and Accountability in Healthcare

The global healthcare ecosystem is composed of a wide range of stakeholders, including governments, health systems, technology companies, implementation partners, insurers, clinicians, and civil society organizations. This ecosystem has been described as one in which “the diffusion of power among state and non-state actors has transformed global health.”¹⁴

In this paper, non-state power refers to the ability of actors outside government to shape consequential decisions that impact healthcare, even when those actors do not hold formal public authority. The influence of non-state actors in healthcare is neither new nor inherently problematic. Companies and other private institutions have long contributed important scientific expertise, resources, products, and services that have helped improve human health and well-being around the world.

At the same time, tensions do arise. The commercial interests of pharmaceutical companies shape research priorities and access, insurers influence which services are covered and under what conditions, and philanthropic organizations shape global health priorities through their funding decisions and technical contributions. Across each of these examples, actors without formal governmental authority exert substantial influence on

technology use and quality of care even as states remain ultimately responsible for regulation, the protection of their citizens, and the stewardship of their health systems.

As a result, non-state healthcare actors have long had to assume certain responsibilities around safety and quality of care. Pharmaceutical companies, for example, generally develop a particular product for a defined medical purpose and bear formal obligations to demonstrate its safety and efficacy. These obligations are supplemented by industry frameworks, such as the International Federation of Pharmaceutical Manufacturers and Associations (IFPMA) Code of Practice, that establish expectations for research, promotion of products, and interactions with healthcare professionals and patient organizations.¹⁵

At the same time, the broader ethical responsibilities for these non-state actors are less clearly defined. Continuing with the example of pharmaceutical companies, considerations such as whether a company should ensure equitable access, prioritize conditions affecting underserved populations, or invest in lower-income markets do not always have comparable regulatory requirements. Rather, these types of obligations are often shaped by voluntary commitments, external incentives, and reputational considerations.

14 Julio Frenk and Suerie Moon, “Governance Challenges in Global Health,” *New England Journal of Medicine* 368, no. 10 (2013): 936–942, <https://doi.org/10.1056/NEJMr1109339>.

15 International Federation of Pharmaceutical Manufacturers & Associations (IFPMA), IFPMA Code of Practice 2019 (Geneva: IFPMA, 2019), https://www.ifpma.org/wp-content/uploads/2018/09/IFPMA_Code_of_Practice_2019_EN.pdf.

For example, frameworks like the UN Special Rapporteur's Human Rights Guidelines for Pharmaceutical Companies address pricing, licensing, and access to medicines, but do not necessarily create enforceable obligations.¹⁶

To that end, there are gaps between what a non-state actor is legally required to do and what it may be ethically expected to do because of the influence its product or services exert over health outcomes. This gap is central to understanding non-state power in healthcare: the more the actor exerts that influence, whether intentional or not, the more important it is to ask what ethical responsibilities should reasonably follow and what are the appropriate frameworks for enforcing or upholding certain behaviors.

Naturally, AI introduces new dimensions to longstanding questions about non-state actors and power dynamics in healthcare. However, not all AI use in healthcare fits the paradigms described above. Some AI systems are developed for defined clinical purposes, such as an AI-enabled diagnostic. These systems may resemble other healthcare products: their intended use can be specified, performance can be evaluated against defined benchmarks and standards, and they can be regulated as a medical device, with obligations¹⁷ attached to the developers and purchasing institutions—even as many of these frameworks and policies are still nascent. AI's self-improving and rapid iteration cycles make oversight and governance a unique challenge compared to more static healthcare products.

General-purpose AI and the agentic systems that society is rapidly moving toward invite a different set of considerations. Broadly capable AI models and systems are not developed as healthcare products nor for singular clinical purposes. Instead, they may be adapted or used by downstream application developers, vendors, health systems, researchers, clinicians, or patients for a wide range of health-related and health-adjacent tasks. While upstream AI model developers shape important features of the AI system such as training and safeguards, the decisions that determine its impact in healthcare are often established by others.

This distinction makes the allocation of responsibility across non-state actors less straightforward. A pharmaceutical company generally knows the condition and regulatory pathway for which a medicine is being developed. A developer of a general-purpose AI model may not know or control every healthcare application built on top of it. Conversely, a downstream developer or health system may determine the clinical use but lack full control over the underlying model. Responsibility therefore cannot rest exclusively with either the upstream developer or the downstream user.

¹⁶ Paul Hunt, "Human Rights Guidelines for Pharmaceutical Companies in Relation to Access to Medicines," annex to The Right of Everyone to the Enjoyment of the Highest Attainable Standard of Physical and Mental Health, UN General Assembly, A/63/263, August 11, 2008, pp. 15–25.

¹⁷ U.S. Food and Drug Administration, "Artificial Intelligence in Software as a Medical Device," content current as of March 25, 2025, <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-software-medical-device>

At the same time, the distinction described above may become less clear over time as general-purpose AI systems advance. A model may not have been designed or authorized to diagnose a disease, and yet as examined in our introduction, it may outperform human clinicians on certain diagnostic tasks. At certain points in performance, ignoring a system's potential use could itself carry ethical consequences if the model could otherwise materially improve accuracy or access to healthcare.

Given these advancing capabilities, stakeholders understandably seek answers as to whether a general-purpose model is capable of performing a clinical task. We position that a separate, but important, set of questions include what evidence, safeguards, and institutional conditions should be required before those capabilities are relied on in the real world. This underscores the need for ethical frameworks to more precisely articulate how accountability is differentiated and distributed across the non-state actors that directly influence healthcare.

Upstream developers are best positioned to evaluate the model's underlying behavior and limitations and provide information that supports responsible downstream use. Application developers, vendors, and health systems may be better positioned to validate a particular healthcare use, establish appropriate human oversight, and monitor performance during implementation. Governments and regulators retain responsibility for setting the legal standards and enforcement mechanisms that govern these relationships and, ultimately, for determining when a broadly capable system crosses into a clinical use case.

Inevitably, more capable AI will put pressure on this distributed model. As systems move from generating information to performing more complex tasks, upstream vs. downstream may blur, and choices about model design and technical safeguards are likely to have more consequential effects. At the same time, decisions by health systems and other end users will continue to shape how those capabilities impact providers, patients, and the relationships between them.

Trust and Operational Accountability

The model of distributed accountability described above has implications for societal trust, both in AI systems broadly and in healthcare. Patient care is already changing as more people use AI both in daily life and in healthcare settings. KFF's Tracking Poll on Health Information and Trust

found that 32 percent of adults had used AI for physical- or mental-health information or advice in the previous year. Among those who had used AI for physical-health advice, 69 percent said they trusted AI tools “a great deal” or “a fair amount” to provide reliable health information.¹⁸

¹⁸ Alex Montero et al., “KFF Tracking Poll on Health Information and Trust: Use of AI for Health Information and Advice,” KFF, March 25, 2026, <https://www.kff.org/public-opinion/kff-tracking-poll-on-health-information-and-trust-use-of-ai-for-health-information-and-advice/>

Notably, the same report found that 65 percent of those who had uploaded personal medical information still reported concerns about the privacy of medical data entered into AI tools. This suggests that trust in AI is not straightforward. Users may trust a system to provide useful or reliable health information while remaining wary of how their data is collected or used. In this sense, trust and confidence in the technology's performance coexists with mistrust of the institutions responsible for developing it, and thus can influence perceptions of accountability.

Trust is also relevant to accountability because patients and clinicians may use the same AI system in different ways. Clinicians may interpret an AI output alongside other medical evidence, professional judgment, and knowledge of the patient, while individuals using a model may treat its response as complete or authoritative. As AI systems become more capable, widely available, and able to perform a broader range of health-related tasks without professional intermediaries, individual differences and preferences across AI use may also become more consequential in shaping accountability.

This underscores the need to operationalize accountability within ethics frameworks. Existing frameworks frequently identify accountability as a core principle, but stating accountability as a value does not establish how it should function in practice. As Donia

observes, “While accountability is often framed as a normative good, less clear is exactly what kind of normative work accountability is expected to do, and how it is expected to do it.”¹⁹ Another unresolved question also centers on how accountability may shift when multiple actors hold responsibility and AI systems become increasingly capable, autonomous, and broadly deployed at a rapid pace.

Donia identifies five distinct “normative logics” of algorithmic accountability in health systems, recognizing that the socio-cultural norms influence how processes and their expected outcomes, including the governance of these processes, are developed. These normative logics of algorithmic accountability include (1) accountability as verification, (2) accountability as participation, (3) accountability as social license, (4) accountability as fiduciary duty, and (5) accountability as compliance. Each corresponds, respectively, to demonstrating that systems perform as expected, involving affected groups, establishing legitimacy and trust, fulfilling professional obligations, and meeting legal requirements.

These logics reflect different assumptions about the purpose of accountability and the actors responsible for carrying it out. In Table 2, we compare Donia’s logics to the ethical frameworks previously summarized in Table 1, highlighting the relationship to trust and operational accountability, as well as the applicable normative logics.

19 Joseph Donia, “Algorithmic Accountabilities and Health Systems: A Review and Sociomaterial Approach,” *Big Data & Society* 12, no. 2 (2025): 1-14, <https://doi.org/10.1177/20539517251334099>

Table 2. Review of Ethical Frameworks for Healthcare AI and Normative Logics of Algorithmic Accountability

<i>Ethical Frameworks for Healthcare AI</i>	Relevance to Themes of Trust and Operational Accountability	Applicable “Normative Logics”
WHO Ethics and Governance of Artificial Intelligence for Health (2021)	Refer to <i>Principle 5.3</i> “ensure transparency, explainability and intelligibility” and <i>Principle 5.4</i> “foster responsibility and accountability.”	Verification, Social License, Fiduciary Duty, Compliance
AMA Trustworthy Augmented Intelligence in Health Care (2024)	Refer to <i>Table 1 Common themes from AI guidelines and regulations</i> , highlighting “accountability—AI systems should be subject to oversight during development and deployment; right remedies should be provided if harm occurs” and “professional responsibility—Individuals and teams involved in the development and deployment of AI systems take responsibility for the performance and effects of those systems.” Refer to <i>Table 2 Crosscutting responsibilities of developers, deployers, and end users in fulfilling commitments to ethics, evidence, and equity</i> , especially highlighting the physician decision-making authority and ability to make clinical decisions such as diagnosis and treatment, and to override the AI system.	Verification, Social License, Fiduciary Duty
WMA Statement on Artificial and Augmented Intelligence in Medical Care (2025)	Refer to the defined “physician-in-the-loop (PITL)” principle recognizing the physician’s authority to “review and retain final authority over all AI outputs before they shape clinical care.” Refer to the <i>Guiding Principles for AI in Healthcare</i> such as “accountability” recognizing that “the prevalence of these tools necessitates clearly distributed accountability . . . among all stakeholders, including but not limited to developers, healthcare organizations, regulators, researchers and clinicians. ”	Verification, Participation, Social License, Fiduciary Duty, Compliance
NAM An Artificial Intelligence Code of Conduct for Health and Medicine (2025)	Refer to <i>Table 3-1 Crosswalk of Draft AICC Code Principles and Commitments</i> , recognizing the applicability of the commitment to “Engage people as partners with agency in every stage of the lifecycle” relevance with the following AICC Code Principles: “engaged,” “equitable,” “accessible,” “transparent,” and “accountable.”	Verification, Participation, Social License, Fiduciary Duty, Compliance

Together, these frameworks establish important foundations for accountability, including professional oversight, informed consent, and patient autonomy. At the same time, they are less clear about how accountability should work in practice—particularly how patients and other affected groups should help shape the design and deployment of AI in healthcare.

As previously described, and with increasingly autonomous AI systems, relying on physician oversight alone will not be enough. There needs to be clearly defined roles, opportunities for participation, and processes for distributing responsibility across many actors. Ethical frameworks must move beyond stating principles to explaining who is responsible for what, how affected people will have a meaningful voice, and what happens when the system causes harm.

Moving from Principle to Practice

Illustrative Example: Clinical Decision Support Systems in Low-Resource Settings

Low-resource healthcare settings provide an important test case for the ethical questions surfaced in this paper. These settings are often characterized by large populations in need of care, shortages of clinicians and specialists, limited diagnostic capacity, constrained budgets, unreliable electricity or Internet connectivity, long distances between patients and facilities, limited infrastructure, and substantial variation in the availability and quality of care.

The World Health Organization (WHO) projects a global shortfall of 11 million health workers by 2030, concentrated primarily in low- and lower-middle-income countries.²⁰ These shortages exist within already challenging healthcare environments: Between 5.7 million and 8.4 million deaths in low- and middle-income countries are attributed to poor-quality healthcare each year.²¹

These conditions may suggest that AI can expand access to expertise and improve the quality of care, but they also heighten the consequences of deploying systems that are not contextually appropriate. AI models can reproduce biases present in their training data, and performance may vary across languages, particularly where medical data in local languages are limited.^{22,23} These risks make local evaluation essential, including assessment across the populations, languages, epidemiology, and existing care pathways in which an AI system will be deployed.

20 World Health Organization, “Health Workforce,” accessed July 17, 2026, <https://www.who.int/health-topics/health-workforce>.

21 World Health Organization, “Quality Health Services,” May 19, 2025, <https://www.who.int/news-room/fact-sheets/detail/quality-health-services>.

22 Stephen R. Pfohl et al., “A Toolbox for Surfacing Health Equity Harms and Biases in Large Language Models,” *Nature Medicine* 30 (2024): 3590–3600, <https://doi.org/10.1038/s41591-024-03258-2>.

23 Samuel Rutunda et al., “Large Language Models for Frontline Healthcare Support in Low-Resource Settings,” *Nature Health* 1 (2026): 191–197, <https://doi.org/10.1038/s44360-025-00038-1>

At the same time, evaluating the trade-offs of AI use in low-resource settings requires an honest assessment of the care patients can currently access. In some settings, the available alternative may be an overextended generalist, prolonged waits for specialist review, or no access to care at all. Recognizing this baseline is not an argument for weaker ethical standards in low-resource healthcare settings: It is a call to action to ensure that AI systems are intentionally designed, contextually appropriate, and assessed against the realities of comparative risk and safeguards needed to realize their potential benefits in these contexts.

Clinical decision support systems (CDSS) illustrate this tension. CDSS use patient-specific clinical information to provide assessments or recommendations that support evidence-based decisions at the point of care.²⁴ AI has expanded these capabilities by enabling systems to analyze larger and more complex forms of health data, including medical images, electronic health records, laboratory information, and clinical text.²⁵ In low-resource settings, these systems have the potential to make clinical knowledge and specialist interpretation available to health workers and patients who would otherwise have limited or no access to them.

Computer-aided detection (CAD) of tuberculosis (TB) on chest X-rays is a powerful example. In settings where radiologists are scarce, these systems can expand TB screening capacity and help identify patients who might otherwise go undetected. WHO now recommends qualifying CAD software as an alternative to human interpretation of digital chest X-rays for TB screening and triage among people aged fifteen and older; in 2025, it approved six software products for this use and described chest X-ray screening with CAD as an essential approach to closing gaps in TB case detection.²⁶

The rapid adoption of AI in radiology suggests that there is a threshold between AI experimentation and routine clinical use that can be and already has been crossed. Radiology accounts for more than three-quarters of AI-enabled medical devices authorized by the U.S. Food and Drug Administration, by far the largest clinical category.^{27,28} A 2024 survey of European Society of Radiology members found that 48 percent of respondents were using AI and another 25 percent planned to do so.²⁹ These figures do not demonstrate even adoption across countries or health systems, but illustrate how purpose-built AI systems can become embedded in routine care once their performance is sufficiently demonstrated and institutions develop workflows for their use.

24 Agency for Healthcare Research and Quality, “Clinical Decision Support Systems,” Patient Safety Network, accessed July 17, 2026, <https://psnet.ahrq.gov/primer/clinical-decision-support-systems>

25 World Health Organization, *Generating Evidence for Artificial Intelligence-Based Medical Devices: A Framework for Training, Validation and Evaluation* (Geneva: World Health Organization, 2021), <https://iris.who.int/handle/10665/349093>

26 World Health Organization, “WHO Approves Six Software Products for Computer-Aided Detection of TB on Chest X-Ray,” June 11, 2025, <https://www.who.int/news/item/11-06-2025-who-approves-six-software-products-for-computer-aided-detection-of-tb-on-chest-x-ray>

27 U.S. Food and Drug Administration, “Artificial Intelligence-Enabled Medical Devices,” accessed July 17, 2026, <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices>

28 Ram Sivakumar, Brian Lue, and Shinjini Kundu, “FDA Approval of Artificial Intelligence and Machine Learning Devices in Radiology: A Systematic Review,” *JAMA Network Open* 8, no. 11 (2025): e2542338, <https://doi.org/10.1001/jamanetworkopen.2025.42338>

29 Moreno Zanardo et al., “Impact of AI on Radiology: A EuroAIM/EuSoMII 2024 Survey among Members of the European Society of Radiology,” *Insights into Imaging* 15 (2024): 240, <https://doi.org/10.1186/s13244-024-01801-w>

“As AI systems become capable of reliably performing certain healthcare tasks, human-centered care may sometimes require less direct human control at certain decision points and stronger institutional accountability around the conditions of use.”

The use of AI in radiology raises forward-looking ethical questions. If an AI system consistently performs a particular task more accurately, and more affordably and accessibly than a human being, continued reliance or insistence solely on human review may not always best serve patients. Under those conditions, the role of the clinician may shift from independently reviewing every image toward supervising the broader diagnostic process, resolving complex or uncertain cases, or helping clinicians and their patients navigate the results.

WHO’s recommendation on CAD for TB screening illustrates the importance of balancing critical questions of access with human accountability. TB is the world’s leading infectious disease killer, sickening over 10.7 million people and claiming over 1.2 million lives annually.³⁰ Access to screening and diagnosis is essential for improving care for millions around the world. In the example of CAD, requiring specialist review of every image would defeat the purpose and public health benefit the technology is designed to help unlock. Instead, people who screen positive must undergo confirmatory testing before treatment, helping maintain human accountability across the broader care pathway.

This example illustrates the broader considerations presented in this paper: As AI systems become capable of reliably performing certain healthcare tasks, human-centered care may sometimes require less direct human control at certain decision points and stronger institutional accountability around the conditions of use. In low-resource settings, that shift may be especially consequential because AI is not just incrementally improving an existing service, it could be delivering a critically important service where there are not enough healthcare workers or resources to provide it.

These questions become more complex as CDSS moves from bounded tools toward general purpose models capable of supporting a wider range of tasks. A general-purpose AI system could assist with history-taking, diagnosis, treatment planning, documentation, referral decisions, and patient communication within a single encounter. And unlike a system developed for one disease or diagnostic task, its role may change across clinicians, facilities, and patient populations so its effects may be more difficult to define and assign to any single actor.

³⁰ World Health Organization, Global Tuberculosis Report 2025 (Geneva: World Health Organization, 2025), “Top Findings and Messages in the 2025 Report.”

In low-resource healthcare settings, there is significant interest in and debate around how CDSS can extend forms of expertise and help improve quality of care. For example, real-world research conducted by Penda Health and OpenAI across nearly 40,000 patient visits found that clinicians with access to an LLM-based clinical copilot made 16 percent fewer diagnostic errors and 13 percent fewer treatment errors than clinicians without access to the system.³¹ Notably, a subsequent cluster-randomized trial provided evidence that while the AI system improved the quality of clinical documentation, diagnostic reasoning, and treatment planning, it did not significantly reduce treatment failure.³² This underscores a set of points we will not explore in depth in this paper, but are worth highlighting: model performance is an incomplete measure of impact, as outcomes also depend on whether patients can obtain medicines, return for follow-up care, and access the services needed to act on those decisions, as well as our ability to measure and evaluate these complex factors in a real-world healthcare system.

Ultimately, the risks of adopting an AI system in any healthcare context must be evaluated carefully, including the risk of error, bias, poor localization, over-reliance, and weak oversight of implementation. But as AI systems become more capable, the evaluation should also account for the clinical context and the potential consequences of non-adoption. When the existing alternative is delayed or entirely absent care, a system shown to perform reliably may help reduce the harms associated with existing gaps in care.

As AI systems become more capable and as we build the evidence base around their use, questions may shift increasingly to include not only whether their use is appropriate, but also the consequences of non-use. This underscores the need for ethical frameworks to evaluate AI against the actual standard of care available in a particular setting, rather than against an idealized clinical environment.

Such frameworks must also recognize that, as AI becomes more capable, accountability must extend beyond whether the model produces an accurate output to whether the state and non-state actors involved have created the conditions required for that output to benefit patients.

31 Robert Korom et al., “AI-Based Clinical Decision Support for Primary Care: A Real-World Study,” arXiv, July 22, 2025, <https://doi.org/10.48550/arXiv.2507.16947>

32 Ambrose Agweyu et al., “Generative AI-Enabled Clinical Decision Support System in Primary Care: A Pragmatic, Cluster-Randomized Trial,” *Nature Medicine* (2026), <https://doi.org/10.1038/s41591-026-04503-6>

Conclusion

Existing ethics frameworks for AI in healthcare provide a strong foundation for governing the next era of AI, with broad agreement that these systems should be safe, equitable, transparent, accountable, and designed to support human well-being. While these principles remain essential, they were largely developed for a world in which AI primarily assists clinicians—not one in which increasingly capable systems operate with greater autonomy, support complex decision-making across care pathways, or interact directly with patients without a professional intermediary.

As AI capabilities advance, the ethical questions will change. Human-centered care should remain the goal, but this does not equate to requiring human oversight of every AI output. Rather, it requires a broader evaluation of how human-centered principles are applied throughout an AI lifecycle. For example, AI-assisted decision-making by physicians could lead to faster and more accurate diagnoses, decrease false reassurance, and improve access to care, particularly in low-resource settings.

Operationalizing accountability in healthcare AI will also require further development. The World Medical Association states that “AI integration does not diminish physician responsibility for patient welfare and advocacy” and that final responsibility for diagnosis and therapy must remain with the physician. At the same time, it maintains that “responsibility should be appropriately allocated among all stakeholders,” including developers, healthcare organizations, regulators, researchers, and clinicians.³³ This

captures a central tension: multiple actors may hold legitimate responsibilities, but those responsibilities will not remain static as the relative influence of those different actors shifts, as highlighted by Donia with the normative logics of accountability.

The unresolved question is how accountability should shift as systems become more capable and assume more consequential roles. Low-resource settings make these questions particularly visible and urgent, especially as there may be additional risks to non-adoption. AI may extend access to diagnostic interpretation and clinical expertise where the alternative is inconsistent or absent care. Realizing that potential, however, requires technologies that are contextually appropriate, locally evaluated, and supported by the infrastructure needed to translate AI outputs into better health outcomes. As systems demonstrate reliable performance, failing to assess or deploy beneficial capabilities may itself contribute to avoidable harm.

The next generation of healthcare AI ethics should review how human oversight and accountability are addressed throughout the AI lifecycle. Ethical frameworks also must be revisited and updated to account for scenarios in which systems are provided permissions to exercise even greater autonomy and capability to enhance patient care. Ultimately, the measure of an ethical framework should be whether it enables increasingly capable AI to improve human health while preserving safety, equity, trust, and meaningful responsibility across the full system of care.

³³ World Medical Association, “WMA Statement on Artificial and Augmented Intelligence in Medical Care,” adopted by the 76th WMA General Assembly, Porto, Portugal, October 2025, para. 9.

Notes and References

Singla, Alex, Alexander Sukharevsky, Bryce Hall, Lareina Yee, Michael Chui, and Tara Balakrishnan. “The State of AI in 2025: Agents, Innovation, and Transformation.” McKinsey & Company, November 5, 2025. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>.

Sidoti, Olivia, and Colleen McClain. “34% of U.S. Adults Have Used ChatGPT, About Double the Share in 2023.” Pew Research Center, June 25, 2025. <https://www.pewresearch.org/short-reads/2025/06/25/34-of-us-adults-have-used-chatgpt-about-double-the-share-in-2023/>.

American Medical Association. 2026 Physician Survey on Augmented Intelligence. Chicago: American Medical Association, Center for Digital Health and AI, March 2026. https://www.ama-assn.org/sites/ama-assn.org/files/2026-03/physician-ai-sentiment-report_0.pdf.

Morley, Jessica, Caio C. V. Machado, Christopher Burr, Josh Cowls, Indra Joshi, Mariarosaria Taddeo, and Luciano Floridi. “The Ethics of AI in Health Care: A Mapping Review.” *Social Science & Medicine* 260 (2020): 113172. <https://doi.org/10.1016/j.socscimed.2020.113172>.

Kung, Tiffany H., et al. “Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models.” *PLOS Digital Health* 2, no. 2 (2023): e0000198. <https://doi.org/10.1371/journal.pdig.0000198>.

Singhal, Karan, et al. “Large Language Models Encode Clinical Knowledge.” *Nature* 620 (2023): 172–180. <https://doi.org/10.1038/s41586-023-06291-2>.

Brodeur, Peter G., et al. “Performance of a Large Language Model on the Reasoning Tasks of a Physician.” *Science* 392, no. 6797 (2026): 524–527. <https://doi.org/10.1126/science.adz4433>.

Kulikowski, Casimir A. “Beginnings of Artificial Intelligence in Medicine.” *Yearbook of Medical Informatics* 28, no. 1 (2019): 263–270.

World Health Organization. Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Geneva: World Health Organization, 2021. <https://www.who.int/publications/i/item/9789240029200>.

American Medical Association. Augmented Intelligence Development, Deployment, and Use in Health Care. Chicago: American Medical Association, November 2024. <https://www.ama-assn.org/system/files/ama-ai-principles.pdf>.

World Medical Association. “WMA Statement on Artificial and Augmented Intelligence in Medical Care.” Adopted by the 76th WMA General Assembly, Porto, Portugal, October 2025. <https://www.wma.net/policies-post/wma-statement-on-artificial-and-augmented-intelligence-in-medical-care/>.

National Academy of Medicine. An Artificial Intelligence Code of Conduct for Health and Medicine: Essential Guidance for Aligned Action. Edited by Laura Adams, Elaine Fontaine, Michael Matheny, and Sunita Krishnan. NAM Special Publication. Washington, DC: National Academies Press, 2025. <https://doi.org/10.17226/29087>.

World Health Organization. Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Geneva: World Health Organization, 2021, 27. <https://www.who.int/publications/i/item/9789240029200>.

Frenk, Julio, and Suerie Moon. “Governance Challenges in Global Health.” *New England Journal of Medicine* 368, no. 10 (2013): 936–942. <https://doi.org/10.1056/NEJMr1109339>.

International Federation of Pharmaceutical Manufacturers & Associations. IFPMA Code of Practice 2019. Geneva: IFPMA, 2019. https://www.ifpma.org/wp-content/uploads/2018/09/IFPMA_Code_of_Practice_2019_EN.pdf.

Hunt, Paul. “Human Rights Guidelines for Pharmaceutical Companies in Relation to Access to Medicines.” Annex to The Right of Everyone to the Enjoyment of the Highest Attainable Standard of Physical and Mental Health. UN General Assembly, A/63/263, August 11, 2008, 15–25.

U.S. Food and Drug Administration. “Artificial Intelligence in Software as a Medical Device.” Content current as of March 25, 2025. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-software-medical-device>.

Montero, Alex, et al. “KFF Tracking Poll on Health Information and Trust: Use of AI for Health Information and Advice.” KFF, March 25, 2026. <https://www.kff.org/public-opinion/kff-tracking-poll-on-health-information-and-trust-use-of-ai-for-health-information-and-advice/>.

Donia, Joseph. “Algorithmic Accountabilities and Health Systems: A Review and Sociomaterial Approach.” *Big Data & Society* 12, no. 2 (2025): 1-14. <https://doi.org/10.1177/20539517251334099>.

World Health Organization. “Health Workforce.” Accessed July 17, 2026. <https://www.who.int/health-topics/health-workforce>.

World Health Organization. “Quality Health Services.” May 19, 2025. <https://www.who.int/news-room/fact-sheets/detail/quality-health-services>.

Pfohl, Stephen R., et al. “A Toolbox for Surfacing Health Equity Harms and Biases in Large Language Models.” *Nature Medicine* 30 (2024): 3590-3600. <https://doi.org/10.1038/s41591-024-03258-2>.

Rutunda, Samuel, et al. “Large Language Models for Frontline Healthcare Support in Low-Resource Settings.” *Nature Health* 1 (2026): 191-197. <https://doi.org/10.1038/s44360-025-00038-1>.

Agency for Healthcare Research and Quality. “Clinical Decision Support Systems.” Patient Safety Network. Accessed July 17, 2026. <https://psnet.ahrq.gov/primer/clinical-decision-support-systems>.

World Health Organization. *Generating Evidence for Artificial Intelligence-Based Medical Devices: A Framework for Training, Validation and Evaluation*. Geneva: World Health Organization, 2021. <https://iris.who.int/handle/10665/349093>.

World Health Organization. “WHO Approves Six Software Products for Computer-Aided Detection of TB on Chest X-Ray.” June 11, 2025. <https://www.who.int/news/item/11-06-2025-who-approves-six-software-products-for-computer-aided-detection-of-tb-on-chest-x-ray>.

U.S. Food and Drug Administration. “Artificial Intelligence-Enabled Medical Devices.” Accessed July 17, 2026. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices>.

Sivakumar, Ram, Brian Lue, and Shinjini Kundu. “FDA Approval of Artificial Intelligence and Machine Learning Devices in Radiology: A Systematic Review.” *JAMA Network Open* 8, no. 11 (2025): e2542338. <https://doi.org/10.1001/jamanetworkopen.2025.42338>.

Zanardo, Moreno, et al. “Impact of AI on Radiology: A EuroAIM/EuSoMII 2024 Survey among Members of the European Society of Radiology.” *Insights into Imaging* 15 (2024): 240. <https://doi.org/10.1186/s13244-024-01801-w>.

World Health Organization. *Global Tuberculosis Report 2025*. Geneva: World Health Organization, 2025. “Top Findings and Messages in the 2025 Report.”

Korom, Robert, et al. “AI-Based Clinical Decision Support for Primary Care: A Real-World Study.” arXiv, July 22, 2025. <https://doi.org/10.48550/arXiv.2507.16947>.

Agweyu, Ambrose, et al. “Generative AI-Enabled Clinical Decision Support System in Primary Care: A Pragmatic, Cluster-Randomized Trial.” *Nature Medicine* (2026). <https://doi.org/10.1038/s41591-026-04503-6>.

World Medical Association. “WMA Statement on Artificial and Augmented Intelligence in Medical Care.” Adopted by the 76th WMA General Assembly, Porto, Portugal, October 2025, para. 9. <https://www.wma.net/policies-post/wma-statement-on-artificial-and-augmented-intelligence-in-medical-care/>.

Their Data, Whose Decision?

The Ethics of Artificial Intelligence in Humanitarian Response

NICHOLAS BAYER

MOLLY SCHAEFFER

ALEX URWIN



CREDIT: Adobe/miss irine

Executive Summary

Artificial intelligence (AI) is rapidly transforming the world and our understanding of it. Across sectors, AI systems are creating significant opportunities while raising complex ethical challenges. This report focuses on one specific dimension of that transformation: the growing integration of AI into humanitarian response.

AI has the potential to improve humanitarian action. It can help responders anticipate crises, allocate resources more efficiently, and reach vulnerable populations faster. But these benefits come with significant ethical risks—not only within the algorithms themselves, but within the data those algorithms rely on and generate.

This report argues that failures in humanitarian AI should not be understood simply as technical breakdowns or problems of model performance. They are often failures of data governance.

Ethical evaluation must therefore move upstream. While questions of model accuracy and effectiveness remain critical—particularly in crisis situations where lives are at stake—greater attention must be paid to how data is collected, who controls it, how long it is retained, and how it may be used in the future.

These questions determine fundamental humanitarian outcomes: Who is visible and who is invisible? Who receives support and who is excluded? And who may face future harm because of data collected during moments of vulnerability?

Through a series of case studies examining humanitarian AI deployments and wider examples of data-driven decision-making—including biometric identity systems, digital service delivery, predictive analytics, and crisis response tools—this report explores how these ethical challenges emerge in practice.

“While questions of model accuracy and effectiveness remain critical—particularly in crisis situations where lives are at stake—greater attention must be paid to how data is collected, who controls it, how long it is retained, and how it may be used in the future.”

Drawing from these examples, we propose a three-part ethical framework:

First, the **minimum necessary technology principle**: Humanitarian actors should use the simplest technological solution capable of addressing the problem at hand, rather than assuming more advanced systems are inherently better.

Second, the **governable data principle**: Organizations should only collect data they can responsibly protect, manage, and govern over time.

Third, the **control and resource principle**: Affected communities must retain meaningful influence over how technologies are deployed and how decisions affecting their lives are made.

This framework is urgently needed. AI systems are moving from experimental pilots into operational deployment across humanitarian organizations, including UN agencies such as United Nations Children’s Fund (UNICEF), United Nations High Commissioner for Refugees (UNHCR), and the World Food Programme. At the same time, humanitarian crises are increasing in scale and complexity, driven by conflict, climate change, displacement, and growing pressure on global systems.

As AI becomes embedded in humanitarian infrastructure, decisions made today may create long-term consequences. The central ethical question is not only whether AI systems work; it is whether the data systems beneath them can be governed responsibly, fairly, and sustainably.

The Benefits of Using AI in Humanitarian Response

A I is already reshaping the way we approach humanitarian responses.

The World Food Programme is using predictive analytics to anticipate food insecurity. Machine learning and satellite imagery are supporting real-time damage assessment following disasters. AI-enabled tools are increasingly being explored for migration management, identity verification, public health modeling, and crisis response.

The potential benefits are substantial. Used effectively, AI can improve the speed, scale, and coordination of humanitarian operations in ways that were previously impossible. Predictive models can help humanitarian agencies anticipate crises before they escalate, enabling earlier interventions that may prevent displacement, food insecurity, or disease outbreaks. During emergencies, automated analysis of satellite imagery, drone footage, and other remote sensing data can provide responders with near real-time situational awareness, helping identify damaged infrastructure, inaccessible communities, and emerging risks while reducing the need to deploy personnel into dangerous environments.

AI also offers important operational advantages. Humanitarian organizations routinely manage enormous volumes of information from governments, UN agencies, NGOs, and local partners. AI systems can synthesize these disparate sources more quickly than human analysts alone, improving coordination across multiple organizations and helping decision-makers prioritize limited resources. Administrative tasks such as translation, document processing, and beneficiary communications can increasingly be automated, allowing humanitarian staff to devote more time to frontline delivery and engagement with affected communities.

These efficiencies are becoming increasingly important as humanitarian needs continue to outpace available resources. Climate change, protracted conflict, and record levels of displacement are placing unprecedented demands on humanitarian systems at the same time that many donor governments are reducing aid budgets. In this environment, technologies that enable organizations to respond more quickly, reach more people, and make better-informed decisions deserve serious consideration.

But the risks are equally significant.

What follows is an examination of those risks through a series of case studies, looking beyond questions of algorithmic performance to consider the underlying data systems that make humanitarian AI possible. These examples demonstrate that many of the most consequential ethical challenges emerge before an AI model is ever deployed: when data is collected, categorized, stored, shared, and governed.

To date, much of the ethical debate around AI—including within humanitarian contexts—has focused on algorithmic bias, transparency, accuracy, and environmental impact. These questions matter. But focusing only on the model or the potential climate impact overlooks deeper ethical questions about safety and doing no harm to individuals caught in crisis.

“Before asking only whether an AI system is accurate or efficient, humanitarian actors must ask whether the data foundations beneath that system are ethical, governable, and accountable over time.”

Why Humanitarian AI Requires a Different Ethical Approach

The ethical questions surrounding AI are not unique to humanitarian response. Across sectors, policymakers and technologists are grappling with issues of bias, transparency, accountability, and governance. However, humanitarian contexts create a distinct ethical environment because of the conditions in which these technologies are deployed. First, humanitarian AI is often used in situations where the individuals affected have limited power. People experiencing conflict, displacement, famine, or natural disasters frequently interact with humanitarian systems because they have no realistic alternatives. This changes the relationship between the institution collecting data and the individual providing it. A person seeking food assistance, shelter, or protection is not engaging with technology under the same conditions as a consumer choosing whether to use a commercial service.

Second, humanitarian decisions often involve exceptionally high stakes. Errors in many AI systems create inconvenience, inefficiency, or financial loss. Errors in humanitarian systems can determine whether someone receives essential support. This does not mean AI should never be used in crisis response. But it does mean that standards for testing, oversight, and accountability must reflect the seriousness of the decisions being made.

Third, humanitarian data is unusually sensitive. Information collected during crises often captures people at moments of extreme vulnerability: where they live, why they fled, what assistance they need, their family relationships, their health status, or their identity. Data collected for protection can become dangerous if accessed by actors seeking to target, exclude, or control those same populations.

Finally, humanitarian crises unfold across changing political environments. Data collected under one government, legal framework, or security situation may exist long after those conditions have changed. The ethical question is therefore not simply: “Is this data useful today?” It is also: “Could this data become harmful tomorrow?” These characteristics mean humanitarian AI requires a higher standard of responsibility. The question should not only be whether a technology improves efficiency. The question must be whether it preserves human dignity, agency, and protection over time. Humanitarians commit themselves to operating with impartiality, independence, and neutrality, but they also commit to protecting human life and alleviating suffering; therefore they agree to do no harm to the people and communities they serve.

Case Studies: Data Risk in Practice

The following real-world examples demonstrate some of the risks of unintentionally harmful data-driven humanitarian systems.

The biometric registration of Rohingya refugees, including data collection by UN agencies later reported to have been shared with the Myanmar government, illustrates how questions of consent and long-term control become critical. Following the mass displacement of more than 700,000 Rohingya from Myanmar to Bangladesh in 2017, humanitarian agencies, including the UNHCR and the World Food Programme, introduced biometric registration systems in refugee camps. Fingerprints and iris scans were collected to improve identity verification, reduce duplication, and ensure aid reached intended recipients. However, in 2021 it emerged that personal data collected by UN agencies had been shared with the Myanmar authorities to support discussions around future repatriation, despite many refugees believing they had not given informed consent for that specific use. Although the intention was administrative rather than punitive, the episode highlighted the fragility of consent in humanitarian settings and the risks created when highly sensitive data outlives its original purpose. The central ethical issue was not whether biometric technology functioned effectively, but whether governance arrangements adequately protected vulnerable people against future political change and secondary uses of their data.

Similarly, the use of digital tools in migration systems demonstrates how personally identifiable data collected under one administration can be used to cause harm later. The U.S. Customs and Border Protection application (CBP One), introduced in 2020 and expanded in 2023, allowed asylum seekers to submit personal information and arrange appointments to present themselves at the U.S. border. For many migrants, use of the application became the only realistic route to accessing the asylum process, meaning participation was effectively mandatory. As U.S. immigration policy changed under subsequent administrations, concerns emerged that information collected to facilitate asylum processing could also support immigration enforcement or removal activities. Whether or not those concerns ultimately materialize, the case illustrates an enduring governance challenge: Humanitarian and migration data are often collected for one legitimate purpose but retained within systems that may later operate under different political priorities. Designing ethical AI therefore requires anticipating not only current uses of data but also future institutional and political change.

India's Aadhaar system demonstrates another challenge: exclusion risk. Launched in 2009, Aadhaar has become the world's largest biometric identity program, enrolling more than 1 billion people through fingerprint, iris, and demographic registration. The system has delivered important benefits by streamlining welfare administration and reducing some forms of fraud, but it has also exposed the consequences of even rare authentication failures. Researchers and civil society organizations have documented cases in which

eligible individuals were unable to access food rations, pensions, or other essential services because biometric authentication failed or supporting infrastructure was unavailable. Although these failures represent a small proportion of total transactions, they become ethically significant when access to basic necessities depends entirely on successful verification. For humanitarian AI, the lesson is clear: Systems that perform well on average may nevertheless produce unacceptable outcomes when the costs of error fall disproportionately on already vulnerable populations. Technical accuracy alone is therefore an insufficient measure of ethical success.

These examples reveal a recurring challenge: Humanitarian technologies are frequently developed in moments of urgency, but the systems they create can persist long after the original emergency has passed.

This creates what might be called a “time horizon problem” in humanitarian AI governance. The immediate incentive during a crisis is to maximize information: collect more data, improve targeting, increase efficiency, and accelerate decision-making. Those objectives are understandable and often necessary.

However, the long-term incentives look different. Data that improves response today may create exposure tomorrow. Information gathered by trusted humanitarian actors may later exist in a very different political environment. A database created for assistance may become valuable for surveillance, migration enforcement, or political control.

This challenge is particularly important because humanitarian organizations often operate in fragile or unstable contexts. The assumption that future institutions will share the values and priorities of current institutions is not always safe. Ethical AI governance therefore requires organizations to consider not only intended use, but possible misuse.

The lesson from these cases is not that data-driven humanitarian response should be avoided. Rather, it is that humanitarian organizations need to consider whether governance models have been designed considering uncertainty. The question is not only “How can this system help?” but also, “What are the risks to our operations and beneficiaries if control over this system, and the data within it, changes?”

Deep Dive on Data Risk

Data sits at the core of humanitarian AI systems. With it comes a distinct set of risks that are amplified by crisis conditions. In humanitarian contexts, these risks are not abstract: they directly influence who receives aid, how people are treated, and what long-term vulnerabilities they may face.

It is important to note that data collection on its own is not necessarily harmful. In fact, having accurate data is among the most important enablers for effective crisis and humanitarian response. Leaders rely on accurate data to determine where needs are the greatest, who is most affected, what resources would be most helpful, and how to prioritize in an environment where timely response is critical.

But it is a delicate balance. Humanitarian ethics emphasize the importance of the “do no harm” principle, and crisis response interventions should never put vulnerable people at greater risk. With this in mind, five interrelated concerns stand out: consent under coercion, biometric permanence, embedded bias, institutional misuse, and unequal control over data. Together, they illustrate how the collection and deployment of data in fragile settings can reinforce existing power imbalances rather than alleviate them. Because each humanitarian emergency is complex and unique, it is difficult to employ a common ethical decision-making framework in different contexts. Even when leaders are aware of the possible implications, ethical trade-offs might be necessary in order to achieve the “best” result.

Consent under coercion is perhaps the most immediate ethical challenge at the point of humanitarian intervention. In crisis situations, the idea of “informed consent” is often compromised by necessity. Refugees and displaced populations entering a refugee camp or health clinic may have little meaningful ability to refuse data collection if doing so risks exclusion from essential services such as food, shelter, or medical care. In these cases, providing consent to a humanitarian or government actor wishing to collect personally identifiable data or medical history could be seen more as a condition for survival rather than a freely given agreement.

People caught in a conflict or political crisis might be distrustful of large international institutions, but ultimately, they will do or say what they must to ensure their safety and basic human rights. Therefore, the risk of coercion, or perceived coercion at least, is quite real. Still, certain information is critical to collect in these settings in order to provide the right medication or legal support. This reveals an ethical tension: collecting data to ensure someone gets timely assistance versus preserving individual agency and freedom over personal information.

In these situations, humanitarian aid workers should have some common-sense solutions on hand to ensure future data protection. Setting rules and parameters around data retention, using statistical disclosure control for critical but sensitive information, and equipping community engagement teams to talk to people about their data and informed consent are all practical actions. In most cases these tactics combined with conditional, time-bound consent to use personal data could be enough to ensure someone’s safety.

In other humanitarian settings, like responses to earthquakes or other natural disasters, sharing personally identifiable data might not be so sensitive to the affected community. Utilizing data or AI systems in these settings might require less scrutiny and could improve speed and efficiency without causing future harm: a win-win scenario. In any context, if the humanitarian responder attempts to collect **biometric data**, concerns about data permanence intensify. Unlike passwords, phone numbers, or identification numbers, biometric identifiers—such as fingerprints or iris scans—cannot be changed or easily revoked if compromised. In humanitarian settings, where individuals may already face persecution or instability, misuse or exposure of this information can create lifelong risks.

Suppose a displaced person entering a camp or clinic is a victim of political, religious, or ethnic persecution. Humanitarians commit themselves to neutrality and recognize the principle of Responsibility to Protect. But they also operate with limited authority and are often only allowed to work in a given context with approval from the governing power in that region. How can humanitarians protect persecuted people if a government demands personally identifiable data for this very group?

In recent history, humanitarians have grappled with shocking requests from governing authorities demanding lists of local partners and beneficiaries participating in humanitarian operations before authorizing their activities. In other contexts, like Gaza and the West Bank, Israeli officials threatened to expel humanitarian organizations who refused to turn over staff lists, suggesting that those who would not cooperate must be complicit in aiding terrorist activity, or at least guilty of abandoning neutrality and impartiality. In these situations, the coercion happens upstream and the humanitarian actors are placed in an impossible position to determine whether to

share personally identifiable or biometric data with authorities at risk of losing access. In the case of Rohingya refugees' biometric data being shared with the government of Myanmar, the UN agencies should have employed a more ethical approach before unconditionally handing over their data.

An important ethical trade-off emerges here: Is it better to risk someone's life today by protecting sensitive data that jeopardizes critical access, or hand over their data to provide critical care today, but risk persecution in the future? Access and lifesaving aid versus compromise, complicity, and harm.

The imbalanced power dynamic between governments and humanitarian actors—and those they serve—raises concerns around **institutional misuse** and **unequal control over data**. Without any parameters, data collected for humanitarian purposes can later be accessed, repurposed, and misused by governments with different legal, regulatory, or political frameworks. Systems initially designed for aid delivery can gradually become tools for surveillance, enforcement, or control, as seen in the example of the CBP One application. Similarly, without parameters around ownership or responsible governance of sensitive data, institutional misuse and unequal control become even more troublesome. Scholars such as Nick Couldry and Ulises A. Mejias have described “data colonialism” as the extraction of data from vulnerable populations by more powerful institutions. As humanitarians consider how to use AI to improve international aid, they must consider whether any data collected from affected communities primarily benefits external actors—through research, innovation, or commercial development—at the expense of, or with limited benefit to the communities themselves.

Humanitarian actors must recognize that maintaining meaningful control over personal data is nearly impossible in today's world, unless one were to go entirely "off the grid." While it might be easy to acknowledge the ethical dilemma of knowingly handing over ultimate control of sensitive data to a government actor, there's also a responsibility for humanitarians to consider who they are doing business with when adopting new AI tools. While those in the private tech sector might best understand the data risks of certain tools, because regulation is often sparse or nonexistent when it comes to new technologies, the short-term responsibility to protect the data of individuals in crisis falls on those coordinating the humanitarian response.

Even when designed with good intentions, technology does not always work, as seen in the case of India's Aadhaar system. Similar examples from health technology also demonstrate the dangers of **embedded bias** in data. Research published in *The New England Journal of Medicine* found that some pulse oximeters produced less accurate readings for darker-skinned patients. When biased data or flawed assumptions are incorporated into automated systems, AI can scale those failures across entire populations, creating a more systemic challenge.

These kinds of technological imperfections raise another ethical dilemma. Humanitarian datasets often reflect existing social, political, and economic inequalities. So if that data is eventually used to train particular AI systems, underlying biases could result in exclusion errors, where vulnerable individuals are denied assistance or allocated decisions that reinforce, rather than reduce, inequity. This result is in direct conflict with the humanitarian commitment to do no harm.

A final consideration on misuse, unequal control, and technological bias relates to the nature of humanitarian response implementation itself. In the current environment, these data risks are not top of mind for individuals implementing crisis responses. Regardless of the context, AI literacy and knowledge of individual data rights and protection are not among the list of competencies required to be a successful humanitarian aid worker.

In moments of crisis, most aid workers are rightly focused on achieving the immediate result of helping as many people as possible, by whatever means necessary—and with little consideration of the ramifications of mishandled data. If these individuals are asked to collect or hand over sensitive data to the authorities, many would comply to achieve short-term benefits, without thinking about longer-term implications for the individuals they are trying to help. In this rapidly changing AI landscape, it is unacceptable and unethical to ignore these risks.

These risks are not hypothetical. As humanitarian organizations increasingly rely on personal data and AI-enabled systems to improve response efficiency, effectiveness, and reach, the leaders in these organizations must employ an ethical decision-making framework to avoid exposing already vulnerable people to new forms of exclusion, surveillance, and harm.

Humanitarian AI systems should minimize unnecessary data collection, avoid irreversible identifiers unless essential, maintain human oversight, and treat data governance as a political and ethical challenge—not merely a technical one.

A Decision Framework for AI in Humanitarian Response

In considering the ethical risks and trade-offs in different humanitarian situations, this report proposes three core questions for leaders to consider before deploying AI in humanitarian response:

1. Should this technology be used at all?
2. Should this data be collected?
3. Who controls the consequences?

Together, these questions form the foundation of an ethical framework for humanitarian AI.

The purpose of this framework is not to slow humanitarian innovation unnecessarily. In many contexts, technological hesitation carries its own ethical cost. Failing to use available tools that could predict famine, improve disaster response, or deliver assistance faster can also result in harm.

The challenge is therefore not choosing between innovation and caution. It is designing a system where innovation happens responsibly.

Each principle below is intended as a practical decision-making tool. Humanitarian organizations and policymakers leading crisis responses should consider these questions before deployment, during implementation, and throughout the lifecycle of any AI-enabled system. Ethical review cannot be a single approval moment before launch; it must continue for as long as the technology and the data behind it continue to exist. Regardless of requirements or regulations, technologists should also keep these ethical considerations in mind when designing and deploying their AI tools and data systems.

QUESTION 1: Should the technology be used at all?

AI has the potential to transform humanitarian response. It can help governments and NGOs anticipate crises, improve coordination, and deliver assistance more effectively. The answer cannot therefore be a blanket rejection of AI.

Instead, this report proposes the **minimum necessary technology principle**: Humanitarian actors should use the simplest tool capable of effectively addressing the problem.

The goal should be functional effectiveness, not technological sophistication.

This principle reflects ideas from medical ethics around the “least invasive intervention.” More complex systems often create additional failure points, reduce transparency, increase maintenance burdens, and make accountability harder.

In humanitarian settings—where resources are limited and affected populations are vulnerable—unnecessary complexity can create unnecessary risk. If the main objective is simple—to provide food and water to anyone in need during a natural disaster—there is probably no need to use an AI system that collects sensitive or personally identifiable data.

The stakes are higher when the affected population is at risk of future persecution, or in need of protections from ill-intentioned actors who could use this technology, and data it collects or stores, to cause further harm. In these humanitarian settings, one should proceed with extra caution and scrutiny. If there are less efficient, though still effective, means to the same end, humanitarians should consider using these tools instead.

The ethical test is therefore simple: Is an AI tool genuinely required, or can the same humanitarian outcome be achieved through a simpler and more accountable approach?

QUESTION 2: Should this data be collected?

AI systems depend on data. The question is therefore not only whether data is useful today, but whether it can be responsibly governed tomorrow.

This report proposes the **governable data principle**: Humanitarian organizations and governments alike should only collect data they can protect, manage, and ultimately delete when necessary.

The ethical test has three questions:

1. Can this dataset be secured?
2. Can this dataset be controlled?
3. Can this dataset be deleted if necessary?

The risks of failing this test are significant. Data may persist long after the original crisis has ended. Biometric data is inherently permanent and may be repurposed for objectives never intended by those who provided it. It may become vulnerable under future political, legal, or security conditions, so it is imperative to only use AI in humanitarian response if institutional misuse of data in the future can be avoided.

QUESTION 3: Who controls the consequences?

The final question concerns power. People affected by humanitarian crises must have transparency, recourse, and meaningful influence over systems that affect their lives.

This report proposes the **control and recourse principle**.

AI systems without appropriate accountability mechanisms risk creating “black box” decisions, where individuals are affected by outcomes they cannot understand, challenge, or change.

This is particularly important in humanitarian contexts, where existing power imbalances between institutions and affected communities are already significant. Humanitarians have a responsibility to mitigate unequal control over data, and data consent under coercion must be avoided at all costs. Technology intended to support vulnerable populations should not reduce their agency.

If sensitive data collected is owned or controlled by a government or a private sector tech company with an ethical compass at odds with humanitarian values, perhaps the system should not be used. As the humanitarian aid sector grapples with these challenges, its leaders may find there is a need for more ownership over sensitive data, and thereby more autonomy to govern and protect it according to humanitarian ethics.

Recommendations for Humanitarian Policymakers and AI Developers

The framework above leads to three practical recommendations.

First, **do not deploy untested AI systems in crisis conditions**. New systems should require validation, independent evaluation, and clear evidence that they improve outcomes before deployment. Embedded bias in the base data of these systems should be avoided or acknowledged and corrected.

Second, **do not collect data without a political threat assessment**. Humanitarian actors should evaluate how data could be misused under future governments, legal frameworks, or security conditions. Where possible, data use should be limited by design.

Third, **ensure accountability and recourse**. Human oversight, appeal mechanisms, and clear institutional responsibility must be built into humanitarian AI systems from the beginning.

Conclusion

The central ethical risks of humanitarian AI are, fundamentally, data governance risks.

AI systems built today may persist for decades. Data collected in moments of crisis can be reused, repurposed, or applied in political and institutional contexts very different from those in which it was originally gathered. Ethical responsibility therefore extends well beyond the point of deployment. It requires sustained attention to how data is collected, governed, shared, retained, and, where appropriate, deleted.

Addressing these challenges will require action across the humanitarian ecosystem.

Governments should strengthen data protection and oversight frameworks for AI deployed in humanitarian settings, ensuring that innovation is matched by robust accountability. Humanitarian organizations and civil society must embed data governance into every stage of program design, while ensuring that affected communities have meaningful transparency, recourse, and a voice in decisions about how their data is used. Data specialists should have a seat at the table to keep protection of vulnerable people's personal data top of mind when building a humanitarian response. Technology developers, meanwhile, should treat humanitarian contexts as requiring the highest standards of data minimization, explainability, and human oversight—not simply another opportunity for AI deployment.

There are encouraging signs that this approach is beginning to take shape. Organizations including the International Committee of the Red Cross and the UN OCHA Centre for Humanitarian Data have increasingly emphasized responsible data stewardship as a core element of humanitarian protection, while international frameworks such as the OECD AI Principles and UNESCO's Recommendation on the Ethics of Artificial Intelligence are helping establish shared expectations around transparency, accountability, and human rights. These initiatives are not a complete solution, but they demonstrate that more ethical models of AI governance are both possible and already emerging.

Responsible humanitarian AI therefore requires not only innovation, but restraint. The best humanitarian technology is not necessarily the most advanced. It is the technology that solves real problems, can be governed responsibly over time, preserves the dignity and agency of affected populations, and minimizes harm.

Ultimately, the future of humanitarian AI will be determined not only by the sophistication of its algorithms, but by the quality of the governance that underpins them.

“The best humanitarian technology is not necessarily the most advanced. It is the technology that solves real problems, can be governed responsibly over time, preserves the dignity and agency of affected populations, and minimizes harm.”

Selected References

Couldry, Nick, and Ulises A. Mejias. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford, CA: Stanford University Press, 2019. <https://www.sup.org/books/sociology/costs-connection>.

Drèze, Jean, Nazar Khalid, Reetika Khera, and Anmol Somanchi. "Aadhaar and Food Security in Jharkhand: Pain without Gain?" *Economic and Political Weekly* 52, no. 50 (2017). <https://www.epw.in/journal/2017/50/special-articles/aadhaar-and-food-security-jharkhand.html>.

GiveDirectly. "Research Evidence on Cash Transfers." Accessed May 15, 2026. <https://www.givedirectly.org/research-on-cash-transfers>.

Haushofer, Johannes, and Jeremy Shapiro. "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya." *The Quarterly Journal of Economics* 131, no. 4 (2016): 1973–2042. <https://doi.org/10.1093/qje/qjw025>.

Human Rights Watch. "UN Shared Rohingya Data Without Informed Consent." June 15, 2021. <https://www.hrw.org/news/2021/06/15/un-shared-rohingya-data-without-informed-consent>.

Internal Displacement Monitoring Centre. 2025 Global Report on Internal Displacement. Geneva: Internal Displacement Monitoring Centre, 2025. <https://www.internal-displacement.org/global-report/grid2025>.

International Committee of the Red Cross. Handbook on Data Protection in Humanitarian Action. 3rd ed. Geneva: International Committee of the Red Cross, 2024. <https://www.icrc.org/en/data-protection-humanitarian-action-handbook>.

Organisation for Economic Co-operation and Development. "OECD AI Principles Overview." Accessed May 15, 2026. <https://oecd.ai/en/ai-principles>.

Organisation for Economic Co-operation and Development. "Recommendation of the Council on Artificial Intelligence." OECD Legal Instruments, OECD/LEGAL/0449. Adopted May 22, 2019; amended May 3, 2024. <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>.

Selbst, Andrew D., Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. "Fairness and Abstraction in Sociotechnical Systems." In FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency, 59–68. New York: Association for Computing Machinery, 2019. <https://doi.org/10.1145/3287560.3287598>.

Sjoding, Michael W., Robert P. Dickson, Theodore J. Iwashyna, Steven E. Gay, and Thomas S. Valley. "Racial Bias in Pulse Oximetry Measurement." *The New England Journal of Medicine* 383, no. 25 (2020): 2477–2478. <https://doi.org/10.1056/NEJMc2029240>.

United Nations High Commissioner for Refugees. Global Trends: Forced Displacement in 2024. Copenhagen: United Nations High Commissioner for Refugees, 2025. <https://www.unhcr.org/global-trends-report-2024>.

World Bank. World Development Report 2016: Digital Dividends. Washington, DC: World Bank, 2016. <https://www.worldbank.org/en/publication/wdr2016>.

World Food Programme. "HungerMap LIVE." Accessed May 15, 2026. <https://hungermap.wfp.org>.

World Food Programme Innovation Accelerator. "HungerMap LIVE." Accessed May 15, 2026. <https://innovation.wfp.org/project/hungermap-live>.

Cohort Overview

Carnegie Ethics Fellows, 2025-2026



The 2025-2026 Carnegie Ethics Fellows cohort. CREDIT: Bryan Goldberg Photography.



JOSHUA ACOSTA

Carnegie Ethics Fellow; Nonresident Peace & Democratic Institutions Fellow, George Mason University

Joshua Acosta is a conflict analysis and resolution expert. His research centers on democratization processes, political economy of security, and alliance politics. Most recently, Acosta has worked as a research associate for the Center for Peacemaking Practice at the Jimmy and Rosalynn Carter School for Peace and Conflict Resolution. As a research associate, he contributed to peacebuilding projects and provided insight as to the frameworks for good governance. He is an affiliate of the Carter School's Narrative Transformation Lab where he provides dialogue regarding the narratives and perspectives of regime types. Acosta graduated from the Jimmy and Rosalynn Carter School for Peace and Conflict Resolution at George Mason University with an MS in conflict analysis and resolution. He also holds an MFA in writing from Randolph College where he was part of a network of authors furthering narratives often overlooked due to political inequality.



NADAV AVIHAY

Carnegie Ethics Fellow; Sustainability Associate, SL Green Realty Corp.

Nadav Avihay has over ten years of experience in sustainable development across government, academia, and the private sector. At SL Green, he leads climate transition and resilience planning and environmental performance reporting. He previously worked at Summit Strategy Group on climate, water, and biodiversity issues, helping companies navigate ESG challenges and responsible business practices, and earlier served as deputy head of the North American branch of Israel's Foreign Investment Authority and IDF's Special Forces. He holds graduate degrees from The New School and Reichman University's School of Sustainability, focusing on the private sector's role in environmental and social objectives.



DENVER BARROWS

Carnegie Ethics Fellow; Senior Product Manager, Amazon

Denver Barrows is a senior manager at Amazon scaling electric vehicle infrastructure. Prior to Amazon, he served in active duty and reserve roles in the United States Air Force with assignments/deployments in Germany, Syria, Puerto Rico, Texas, and California. He continues to serve in the U.S. Air Force Reserves. Barrows is an Atlantic Council Veteran Advanced Energy Project Fellow, has been recognized as a Center for New American Security Next Generation National Security Fellow and on National Security's Out Leadership List, and is an alum of Globally's Emerging as a Global Leader Experience. He was an AFROTC Distinguished Graduate at The University of Chicago where he earned a BA in public policy. Barrows earned his MsC in logistics from the Air Force Institute of Technology where his research focused on how FEMA could implement improved methodologies for selecting transportation service providers during disaster relief events.



NICHOLAS BAYER

Carnegie Ethics Fellow; Senior Communications Manager for Public Engagement, Doctors Without Borders/Médecins Sans Frontières (MSF)

Nicholas Bayer is the senior communications manager for public engagement at Doctors Without Borders/Médecins Sans Frontières (MSF). Since joining the organization in 2017, Bayer has coordinated and supported hundreds of events, from discussion series and virtual project visits to large scale multimedia exhibitions and global conferences. As a member of the communications department leadership team, he upholds MSF's core mission to bear witness and elevate the voices of the patients in the communities where MSF provides medical care. Prior to joining MSF, Bayer held positions at HuffPost Pollster, the Marist Institute for Public Opinion, and The British House of Commons. He also contracts with the Times Square Alliance on event production work for key events like the New Year's Eve ball drop. Bayer graduated from Marist College with a degree in political science and international affairs, with minors in French and music.



REBECCA DISTLER

Carnegie Ethics Fellow

Rebecca Distler is an impact-driven leader with deep expertise at the intersection of AI and healthcare. She is a member of the Strategy team at OpenAI where she works to help translate frontier AI capabilities into public benefit through partnerships and strategic initiatives. Throughout her career, Distler has worked across technology companies, foundations, nonprofits, and governments to help ensure emerging technologies benefit the communities that stand to gain the most. Prior to OpenAI, she built and led the AI and Digital Health portfolio at the Patrick J. McGovern Foundation, a philanthropy advancing AI for public good, and served as director of Global Health Initiatives at Element, a deep learning company co-founded by Yann LeCun. Distler holds a Master's in health policy and global health from the Yale School of Public Health and a BA in political science from Yale University. She is a term member at the Council on Foreign Relations, an AI XPRIZE semifinalist, and an alumni World Economic Forum Global Shaper.



KATHY KIM

Carnegie Ethics Fellow

Kathy Kim is a mission-oriented leader engaging in AI, data, innovation, and partnership-building across public and private sectors, working with c-suite and cross-functional leaders. Through her leadership and volunteerism, Kim has been recognized as a “Rising Star” by Women in Technology, fellow with the Atlantik Bruecke Young Leaders Conference in Germany, and Global Shaper for the World Economic Forum Annual Meeting of New Champions in China. Previously, she led on the development of an enterprise AI governance system at Booz Allen Hamilton, and also led public sector engagement on AI strategy, applied data science, and technology partnerships for federal civilian and intelligence community agencies. Currently, Kim is at the Executive Office of the President Office of Management and Budget’s Office of the Federal Chief CIO, where she engages chief AI officers across the federal government and contributes to government-wide technology policy, including responsible AI adoption and innovation. Her work shapes interagency collaboration and policy guidance designed to ensure AI systems are deployed safely, effectively, and in ways that benefit the public.



RUTH NASHIPAE MUIGAI

Carnegie Ethics Fellow; Founder & CEO, The Gender Initiative

Ruth Nashipae Muigai is founder and CEO of The Gender Initiative, leading gender equality and inclusion efforts in East Africa, with degrees in international law and human rights from the University of East London. A recipient of the Commonwealth Scholarship, Mandela Washington Fellowship, and Transform Norms Fellowship, she is a published researcher on women’s rights, climate justice, and transitional justice, recognized as an African climate expert by AGNES and a certified UN climate negotiator contributing to YOUNGO’s Women and Gender Working Group.



MOLLY SCHAEFFER

Carnegie Ethics Fellow; Deputy Commissioner of Policy and Strategy, New York City Department of Correction

Molly Schaeffer has served as deputy commissioner of policy and strategy at the New York City Department of Correction since July 2026. In this role, she is helping shape the future of the Department and New York City’s criminal justice system, including building new jails and the work required to close Rikers Island. Prior to this position, Schaeffer served as the founding executive director of the Office of Asylum Seeker Operations (OASO), where she led the city’s response to the migrant humanitarian emergency. She began her career in city government in the Chancellor’s Office at NYC Public Schools and later served as an advisor at the NYC Department of Health and Mental Hygiene. Schaeffer is a graduate of the University of Southern California and a lifelong New Yorker. She completed the Executive Leaders Program at the Naval Postgraduate School’s Center for Homeland Defense and Security in 2023.



HARSH SURI

Carnegie Ethics Fellow; Co-founder & CEO, The Geostrata; Founder, STRATA GPT

Harsh Suri is co-founder and CEO of The Geostrata, one of India's leading youth-led think tanks focused on global affairs, diplomacy, climate action, and strategic policy, and founder of STRATA GPT, an AI-driven initiative leveraging emerging technologies for policy, research, and communications. He has worked extensively on international dialogue, youth engagement, and strategic communications, organizing the NATO-India Youth Conference and representing youth voices in global policy discussions.



ALEX URWIN

Carnegie Ethics Fellow; Head of Strategic Partnerships and Projects, UK Prime Minister's Office

Alex Urwin is head of strategic partnerships and projects in the UK prime minister's office. This is his second stint in Number 10, having previously worked in the prime minister's research team. He's also worked as special advisor to the CEO at the Children's Investment Fund Foundation, one of the world's largest climate and development philanthropies. In addition, Urwin has served as an advisor to the CEO of the inaugural Africa Climate Summit, and as a speechwriter and advisor to the COP26 president. He has contributed commentary and analysis to the likes of Foreign Policy, Atlantic Council, and Climate Home, as well as curating and participating in podcasts for Aspen UK. Outside of the world of foreign policy, he's also a published non-fiction author, award-winning playwright, and aspiring screenwriter. Urwin has degrees from Oxford and Columbia, and he's also an Aspen UK Rising Leaders Fellow.

Carnegie Council for Ethics in International Affairs is an independent and nonpartisan nonprofit. The views expressed within this project are those of the authors and do not necessarily reflect the position of Carnegie Council or the authors' employers.

THE CARNEGIE ETHICS FELLOWSHIP

In today's world of geopolitical upheaval and global economic transformation, where can young leaders go to understand the power of ethical leadership, its impact on multilateral cooperation and collaboration, and how it applies to their professional and personal lives?

The Carnegie Ethics Fellowship is a space for talented young professionals to develop their capabilities and be examples of values-driven responsible leadership. Fellows collaborate on projects curated by Carnegie Council for Ethics in International Affairs, giving them the opportunity to contribute to work that has deep connections to both New York and the broader world. The two-year Fellowship is structured to develop the next generation of ethical leaders from business, government, academia, and nongovernmental organizations. The Fellowship is part of Carnegie Council's significant commitment to developing ethics in leadership and to the communities of experts that work toward this end, aligning the power of decision-making with reflective right action.

ABOUT CARNEGIE COUNCIL

Carnegie Council for Ethics in International Affairs is an independent nonprofit that works to empower ethics globally by identifying and addressing the most critical ethical issues of today and tomorrow. Founded by Andrew Carnegie over a century ago, we set the global ethical agenda and work for an ethical future by convening leading experts, building active communities, producing agenda-setting resources, and catalyzing the creation of ethical solutions to global problems. Join us in using the power of ethics to build a better world. Carnegie Council is a nonprofit 501(c) (3) institution. For more information, please visit CarnegieCouncil.org