INSIDE: **Big ideas** from Rumman Chowdhury, Jonathan Julion, Camille Stewart Gloster, Hannah Ismael, Chinasa T. Okolo, Mlindi Mashologu, Martina Le Gall Maláková, Paola Galvez-Callirgos, Louis-Victor de Franssu, Theodoros Evgeniou, Michelle Daniel and Thiago Guimaraes Moraes

FEBRUARY 2025

AI ACTION SUMMIT EDITION

Compiler

TECH'S NEW COLD WAR

The fight for AI dominance is pushing global cooperation to the brink, and the consequences could be catastrophic.

Tekendra Parmar

In November 2023, leaders from 28 countries—including the United States, China, India, Nigeria and representatives from the European Union—gathered at Bletchley Park in Buckinghamshire, England, to sign a historic declaration of cooperation. Echoing treaties from the Atomic Age, the Bletchley Declaration affirmed its signatories' commitment to design, develop and regulate artificial intelligence while ensuring the safety of humankind. Days before the summit, then-British Prime Minister Rishi Sunak warned that the existential risk of AI was akin to the threat of nuclear war. To confront these potential societal harms, he announced the world's first AI Safety Institute. Shortly thereafter, the Biden administration issued an executive order establishing an American AI Safety Institute while advocating for responsible innovation that safeguards workers' rights. But this collective effort is quickly unraveling. Instead of gathering global cooperation, we are now engaged in a tête-à-tête arms race between AI superpowers.

CONTINUED ON PAGE4

Compiler is a 501(c)(3) nonprofit newsroom that launched in 2024 to report on the people, institutions and global forces shaping our digital future. It delivers expert journalism, ideas and research to better inform critical decisions about the role of technology in society.

Compiler aims to expand access to policy news and elevate diverse voices and viewpoints all too often left out of the conversation. This is a publication for and about people making tech policy or working to influence it—from anywhere in the world. Our journalism is independent, free, solution-oriented and global—and available online, in person and in print.

FOUNDER AND CEO
Mike Farrell

BOARD MEMBERS
David Grant Melanie Teplinsky

CONTRIBUTING EDITOR
Tekendra Parmar

ART DIRECTOR
Matthew Curry

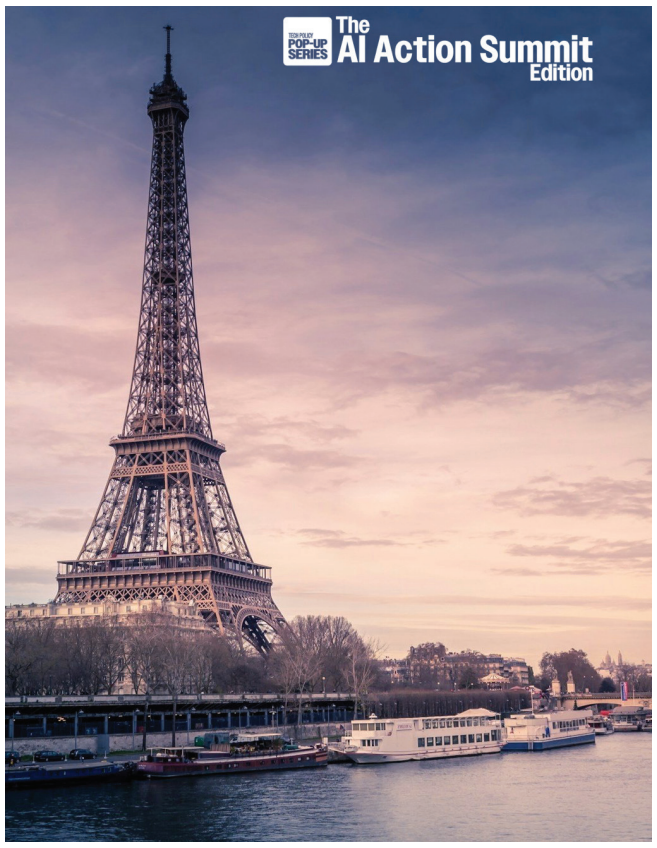
COPY EDITOR
Alison Aves

CONTRIBUTORS
Pallavi Bhargava
Ellery Biddle
Andrew Bochman
Alexander Botting
Brian Calvert
Dave DeWalt
Mlindi Mashologu
Issie Lapowsky
Toni Lorente
Nicolas Moës
Malka Older
Lorena O'Neil
Eric Pickersgill
Philip Reitinger
Alexandra Samuel
Nancy Scola
Aliya Sternstein
Shaun Waterman

LOGO AND BRANDING
Upstatement

CONTACT
COMPILER MEDIA, INC.
PO Box 21095 Washington, D.C. 20009
info@compiler.news

Introducing the AI Action Summit Edition



Artificial intelligence is the last thing I associate with Paris. Like most, I think of food, wine, fashion, art or even the booksellers along the Seine. But AI has arrived in a big way. The French government's AI Action Summit is bringing together world leaders, CEOs, technologists, advocates, civil society and journalists to debate the profound changes AI is driving worldwide. There will be talk of innovation—how automation is reshaping health care, education and work. Some will warn of overregulation; others will sound the alarm on unchecked corporate power and existential risk. But too often, Global Majority voices are absent from these discussions.

This special AI Action Summit Edition of Compiler is meant to amplify those perspectives. As Camille Stewart Gloster, former White House Deputy National Cyber Director, puts it: "To foster equitable global innovation in artificial intelligence, every country—no matter its size or location—needs a way to participate." Camille and I collaborated to bring this issue—and an accompanying pop-up event on the sidelines of the summit—to life, centered on Global Majority concerns in AI.

For this edition, we invited experts, innovators and civil society leaders to contribute their ideas for the responsible development of AI in Global Majority countries. I hope this issue, like all of Compiler's work, helps to rebalance the global tech policy conversation.

— Mike Farrell, Founder & CEO, Compiler Media, Inc.

SIGN UP FOR THE LATEST UPDATES AT www.compiler.news

Trustworthy Innovation for a Connected World.

February 11, 2025 | Paris, France

This action-oriented gathering will focus on how Global Majority countries are building responsible and safe AI-enabled futures. It will bring together regional leaders, technical experts and industry innovators to discuss next-generation tools, practical solutions and new approaches to creating trustworthy, self-sustaining AI ecosystems that address global and local challenges. Building on insights from the 2024 Global Action Forum and the report "The Global Majority AI Agenda: The Path to Shared Prosperity Is Anchored in Equity and Sustainability," the event will explore the intersection of trust, safety and AI, showcasing both opportunities and potential risks for online safety.

"The ideal of shared prosperity in the global human community is undermined by entrenched power imbalances and the absence of a large number of countries from critical decision-making processes and fora. For there to be genuine global collaboration, global leaders across industry and government must be intentional in redressing lack of inclusivity and inequities across the AI tech stack as well as governance mechanisms. Humanity will be much better off if the existing and coming AI wave lifts all nations rather than the alternative of smaller, less wealthy nations being left to compete on an uneven playing field with rules that are written by and unjustifiably favor larger, wealthier, more powerful nations."

– The Global Majority AI Agenda: The Path to Shared Prosperity Is Anchored in Equity and Sustainability

FEATURED SPEAKERS

ANDREW WILSON
Deputy Secretary General (Policy), International Chamber of Commerce (ICC)

CAMILLE STEWART GLOSTER
CEO and Principal, CAS Strategies, LLC, former Deputy National Cyber Director, The White House

CHINASA T. OKOLO
Fellow, Center for Technology Innovation, Governance Studies, The Brookings Institution

CORDEL GREEN
Executive Director, Broadcasting Commission of Jamaica, and Convenor, Caribbean AI Initiative

LOUIS-VICTOR DE FRANSSU
Cofounder, Tremau

MICHELE LAWRENCE JAWANDO
Senior Vice President, Omidyar Network

MIKE FARRELL
Founder and CEO, Compiler Media, Inc.

MLINDI MASHOLOGU
Deputy Director-General (DDG), Digital Society and Economy, National Department of Communications and Digital Technologies, South Africa

NARUN POPATTANACHAI
Director of Impact Assessment and Evaluation of Law, Law Reform Division, Office of the Council of State, Thailand (OCS)

PAOLA GÁLVEZ-CALLIRGOS
Technology Policy Adviser, MPP University of Oxford

RUMMAN CHOWDHURY
CEO and Cofounder, Humane Intelligence

SESSION 1
ADVANCING MULTISTAKEHOLDER GOVERNANCE OF AI

SESSION 2
A NEW ERA OF TRUST & SAFETY?

SESSION 3
CATALYZING AI INNOVATION ECOSYSTEMS WORLDWIDE

HOSTED BY

CAS Strategies · Compiler · OMIDYAR NETWORK™ · KNIGHT FOUNDATION · TREMAU

LEARN MORE AT WWW.CAS-STRATEGIES.COM/AI-ACTION-SUMMIT

OFFICIAL EVENT OF SOMMET POUR L'ACTION SUR L'IA

Tech's new cold war

CONTINUED FROM COVER

As a result, the Doomsday Clock, a signal for how close humanity is to annihilation—made by the very scientist who worked on the nuclear bomb—has moved forward one second, to 89 seconds to midnight. We are, at least by this calculation, aggressively moving ourselves toward extinction. One of the chief concerns among the clock keepers, the Bulletin for Atomic Scientists, is the unchecked rise of AI..

This special issue of Compiler, much like this week's AI Action Summit in Paris, is an effort to reframe the conversation away from wanton corporate competition among global adversaries. In it you will find opinions, analysis and features that parse how we got here while presenting a vision for a sustainable AI future.

That goal has become even more challenging. Last month, the new US presidential administration repealed Biden's executive orders on AI just hours after Trump took office. In its place, the administration set forth a new vision for American AI policy—one in which innovation is unfettered by regulatory burdens. Similarly, Trump's FTC chair, Andrew Ferguson, said AI regulators are on the wrong side of the debate. This anti-regulatory stance, according to the new administration, will help America win the AI race—but at what cost? The last decade has foreshadowed what unchecked technological growth may impose on society. As Biden warned of the growing power of the tech-industrial complex, these firms have latched on to the renewed jingoism. A week before Trump's inauguration, OpenAI released a 15-page memo expressing its views on how the United States can beat China in the AI race. If Washington doesn't act, it warned, AI investments will go to China-backed projects, "strengthening the Chinese Communist Party's global influence." The firm is taking a page from former Google CEO Eric Schmidt's "The New Digital Age," outlining a new era in which Silicon Valley is the engine of America's global influence.

As the Trump administration rejects AI regulation, the appetite in the EU for curtailing the power of Big Tech is growing. In March 2024, the European Union passed the AI Act, the most comprehensive piece of AI regulation to date. The act takes a risk-based approach to regulation, banning certain activities, such as using algorithms for social scoring, while imposing stringent rules on AI in national infrastructure, legal systems and other critical areas. After the passage of these regulations, OpenAI CEO Sam Altman—who once warned of the existential risks posed by the very systems he was building—threatened to leave the EU if compliance with its regulations became too cumbersome. The company was at least somewhat successful in lobbying to dilute some of the act's terms, according to reporting by Time magazine.

But Europe isn't a unified bloc when it comes to countering Big Tech. France, the host of the AI Action Summit, lobbied alongside American firms to weaken portions of the AI Act. The country is home to Mistral AI, one of Europe's biggest AI firms. France's position is that stringent regulations may prevent the creation of a European alternative in the face of Chinese and American dominance. "Rather than lamenting that the great digital champions are America today and China tomorrow, let us put ourselves in a position to create European champions," French president Emmanuel Macron said in a recent speech at Sorbonne University.

Alongside the West's unsettled debate over regulation and how to compete in the AI race, China's technological prowess continues to grow. A previously little-known Chinese startup called DeepSeek has suddenly disrupted the Western AI industry by releasing a model as sophisticated as OpenAI's latest, but trained at a fraction of the cost. The model, which cost only $6 million to make, wiped $1 trillion off the combined value of U.S. tech stocks.

Technologists and policymakers have reacted to this new AI contender with a combination of fear, envy and admiration. "DeepSeek-R1 is AI's Sputnik moment," venture capitalist Marc Andreessen wrote on X. But the moment didn't come out of nowhere. China is the largest producer of AI research in the world. At the same time, it has some of the most robust AI regulations that outline disclosure requirements, model-auditing mechanisms and technical performance standards. Of course, many of these regulations are in service of China's strict censorship regime: Ask DeepSeek's R1 about Tiananmen Square and it will tell you that the question is beyond its programming. Question it about the sovereignty and ownership of Taiwan or the Spratly Islands in the South China Sea, and predictably, R1 forcefully stakes China's claim to the lands.

Despite this apparent censorship, R1 is a soft power coup. Not only will DeepSeek make efficient AI models more accessible to the global market, including in the West, but the company also made its training process public. If the reported costs around DeepSeek's creation are accurate, and there is evidence to suggest that they aren't, it may inspire AI development in economies that previously believed themselves priced out of the market. In our current global trajectory, it is, of course, likely that competition, barriers to entry and the power of data will subjugate the former colonized vassals once again.

Tech companies have never been accountable to users in the Global Majority. Meta once promised to bring the benefits of the internet to billions of users across Africa, Asia and Latin America, only to be accused of facilitating ethnic conflicts everywhere from Myanmar to Ethiopia due to unchecked disinformation and misinformation on its platforms. When a court in Kenya tried to hold Meta accountable for its actions, the company argued that the African court held no jurisdiction over the American company. Given this precedent, it is hard to believe that the companies building powerful AI will voluntarily heed the interests of the most vulnerable. It also underscores the importance of a strong Global Majority coalition to help lead AI regulation.

India, the co-chair of the AI Action Summit, is a prime illustration of the dangers AI poses to emerging economies. AI threatens to automate two backbones of the country: outsourcing and manufacturing. The IT industry in India employs nearly 6 million people performing routine programming and data management. Over half of those workers fear losing their jobs to automation within the next five years, according to findings in the country's latest economic survey. As factories shift toward automation, India risks both job losses and the reshoring of production, with wages for those who remain employed likely to be pushed even lower as they compete with robots at home and abroad.

Yet initiatives to retrain India's workforce for the AI era have been slow to evolve. The Indian billionaire Nandan Nilekani—known as the cofounder of Infosys as well as the chief architect of Aadhaar, India's largest digitizing initiative—advertised his foundation's Springboard program as one such retraining effort in Time magazine's AI 100 list. It claims to have over 400,000 learners, but it's unclear whether this initiative can be scaled to meet the volume of AI-related job losses.

Instead of global cooperation, we are now engaged in an tête-à-tête arms race between AI superpowers. As a result, the Doomsday Clock, a signal for how close humanity is to annihilation—made by the very scientist who worked on the nuclear bomb—has moved to 89 seconds to midnight.

Nilekani has recently advocated that, unlike China, India should temper its dreams of entering the AI race and building its own LLM. Instead, it should focus on building data centers and applications that use pre-existing models. The advice echoes one of Silicon

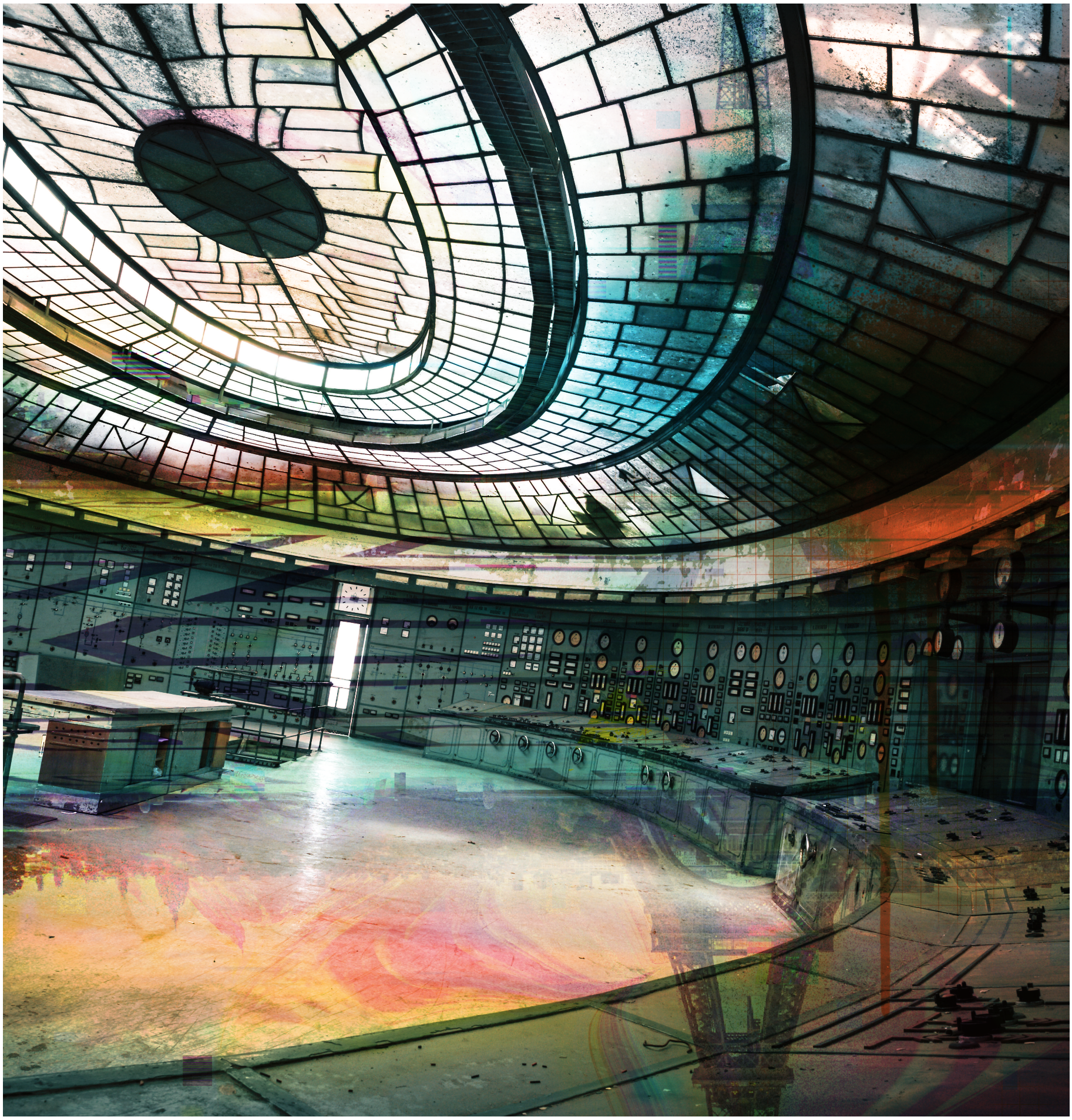
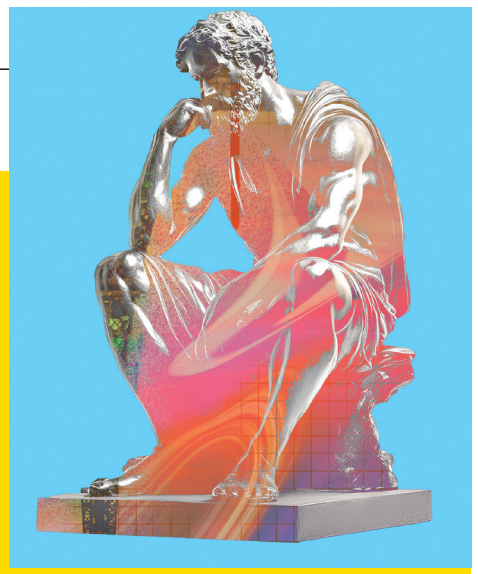


Photo illustration Matthew Curry

Valley's key talking points to the economically impoverished: Data centers will bring growth and employment to the cities and towns left behind by decades of economic reorientation and offshoring. But from Milwaukee, Wisconsin, to Jamnagar, Gujarat, the jobs produced by constructing these centers will be tenuous at best and will not sustain employment like the traditional manufacturing industries of the past.

India is only one example of how much labor retraining efforts among the Global Majority are lagging. But it is a bellwether for those who will listen. A failure to focus on retraining efforts will aggravate pre-existing issues of civil unrest within developing economies and the resulting mass migrations to developed ones.

From the events of the last year, it is evident that the cynicism of global competition has put national interest at odds with the safety and security of humanity. As tech companies and their host nations vie to outcompete each other, AI is already proliferating across our lives. By this year alone, the World Economic Forum estimates AI will displace 85 million jobs globally. The knock-on effect of these disruptions will not be confined within geographical borders. Paris is an opportunity to redirect the conversation, and for those willing to do so, Compiler is a blueprint.



We need to rethink trade secrecy to build better AI

Trade secrecy isn't just about keeping AI models under wraps—it actively encourages secrecy, stifles competition, and limits innovation.

Hannah Ismael

The release of OpenAI's GPT-4 in the spring of 2023 came with a curious disclaimer in its technical report: "[T]his report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar." The technical information, previously made open—hence the company's name—was now made secret. The New York Times filed a lawsuit against the company that called out the decision, but OpenAI's chief scientist defended it as "secrecy on commercial grounds."

The opportunity to decide how a model gets built, especially a tool used as frequently and widely as generative AI, is a position of great power. But trade secrecy closes off that power to all but a few individuals. It also makes the tool a black box. Allowing AI experts and civil society to pry into that black box can ensure that models are more representative of diverse perspectives and help catch and address potential harms AI might pose to society.

Trade secrecy, interestingly enough, not only allows for secrecy but, through the matter of its legal requirements, encourages it. The laws around it have a fundamental requirement that "reasonable efforts" be taken to prevent the release of the trade secret. This compounds the effects of the secrecy regime.

Noncompetes and NDAs are measures frequently taken to achieve the end of reasonable efforts, but they also prevent the free movement and flow of already limited technical expertise. Moreover, it's hard to ascertain whether a trade secret is actually worthy of that distinction without publicly disclosing it. This is, as University of Chicago legal scholars have called it, "trade secrecy's information paradox."

Furthermore, IP and trade secrecy can actually harm the very innovation they aim to foster by gatekeeping for a landscape with only a few players. In the case of AI, this can impact model quality down the line. For example, Stable Diffusion, the image-generation model, often produces a distorted version of the Getty watermark, presumably after being trained on watermarked photos (a practice for which it is being sued). Applications built on top of this model or companies that integrate it into their workflow are at risk of reproducing the error. The error is an example of how algorithms in a consolidated market can produce erroneous results that then become embedded and amplified downstream.

One way to address the high barriers to entry and the concentration of the market is by encouraging global investment in public infrastructure across the AI supply chain. Given the purposes AI serves for the public, it would be reasonable for individual countries to invest in publicly accessible hardware, software and data—or Public AI—for open-source organizations. This can exist as grants provided by governments to organizations seeking to democratize access to these resources (such as by supporting individuals who are creating cheaper proprietary datasets) or to individuals seeking to create an open model and needing funding to access downstream resources. The UN has already carved out funding for this purpose in its Governing AI for Humanity Report, though whether it will actually be implemented remains to be seen.

However, resolving market concentration doesn't fix AI's black box problem; only a systemic shift within our approach to disclosure can achieve this. European transparency legislation mandates that AI companies produce documentation describing how their models are trained, how they function and what risks they pose. This sort of partial openness forces companies to develop records of their information in a way that allows civil society an opportunity to decide whether or how to investigate harms. It also offers regulators greater clarity in understanding whether claims are valid. The act aims to balance transparency and intellectual property, allowing companies to document how their models work without truly revealing the "secret sauce." Of course, transparency documentation has its own flaws, namely that it may produce another opportunity for firms to self-govern. However, it acts as a starting point for legislators to reconsider: Is secrecy really achieving what it set out to do?

Hannah Ismael works on global AI policy at Mozilla.

A blueprint for a new and equitable global data trade

Data-trading alliances would democratize access to critical resources, foster collaboration and accelerate innovation in areas like AI, health care and sustainability.

Camille Stewart Gloster

In order to foster equitable global innovation in artificial intelligence, every country—no matter its size or location—needs a way to participate. That is why it's imperative to create a new and accessible global trade in data, the essential ingredient for any advancement in AI. A global data market will not just accelerate innovation but create opportunities for progress in areas such as health care, climate-change mitigation, international security and beyond.

At the moment, most countries are grappling with whether to treat data like air or oil. Treating it like oil recognizes the commoditization of data; treating it like air frames it as a ubiquitous resource vital for collective well-being that should therefore be freely shared and collaboratively used for the common good.

In practice, many countries find themselves balancing the two perspectives—seeking to maximize the economic potential of data while recognizing the societal benefits of treating it as a shared resource. The key lies in developing frameworks that respect data sovereignty and privacy while encouraging collaboration and innovation.

A new model for global digital trade would enable countries to use data as a valuable asset to barter for necessary resources such as technology, energy and other commodities. To preserve data sovereignty and privacy, this system could operate through a secure global digital marketplace to which countries contribute anonymized datasets. The value of the data would be determined using standardized metrics such as quality, volume and potential for generating insights or driving innovation in specific areas. Transactions could be facilitated using blockchain for transparency, and smart contracts would enforce terms of use and protect against misuse.

When multiple entities collaborate to train AI models by sharing data, it's known as federated learning. Federated learning can play a pivotal role in the global-market model by enabling countries to share the value of their data without transferring the raw data itself—particularly important for highly sensitive datasets. In this system, algorithms would analyze a country's local data, and only aggregated, anonymized insights or model updates would be shared. This ensures that sensitive information remains secure and compliant with local regulations, while still allowing everyone to contribute to and benefit from global advancements. For example, a country with rich biodiversity data could use federated learning to train global AI models for climate adaptation, and in return, gain access to technology or funding to protect its ecosystems. This would

allow us to advance collective sustainable development goals while acknowledging the value of the data and protecting the interests of the country and its citizens.

Importantly, this system would prioritize inclusivity. It would enable smaller or resource-scarce nations to participate meaningfully by offering high-value datasets in areas like renewable energy potential. Such a system would not only democratize access to critical resources but also foster global collaboration, with shared data driving innovations. Equally important, models like this allow for security, privacy and other safety principles to be part of the negotiations for data use and governance. By treating data as a strategic asset, this model could redefine international trade and promote equitable economic growth.

International laws would need to adapt to enable data to be bartered for necessary resources as part of digital trade agreements. Current frameworks governing data—such as privacy regulations, intellectual property laws and international trade agreements—are not designed to treat data as a tradable commodity on par with traditional goods and services. For instance, privacy laws, like the EU's General Data Protection Regulation or data localization mandates in countries such as India and China, impose restrictions on how data can be transferred or used across borders.

Additionally, international trade agreements, including those governed by the World Trade Organization, lack specific provisions for data as a barterable resource. Defining data as a tradable asset within these frameworks would require legal classifications and mechanisms for dispute resolution, taxation and valuation. Intellectual property laws would also need to be updated. Harmonizing these legal changes across jurisdictions would be critical to ensuring fairness, trust and accountability in a global data barter system, allowing countries to exchange data for resources while addressing concerns related to privacy, security and equitable access.

The work to form new digital trade agreements can begin now through bilateral or regional trade agreements. Countries can initiate pilot programs to showcase the feasibility and value of data barter. For example, one country might partner with private-sector entities or international organizations to create a secure and ethical data-sharing model. These initiatives could serve as proof of concept, strengthening a nation's position in negotiations.

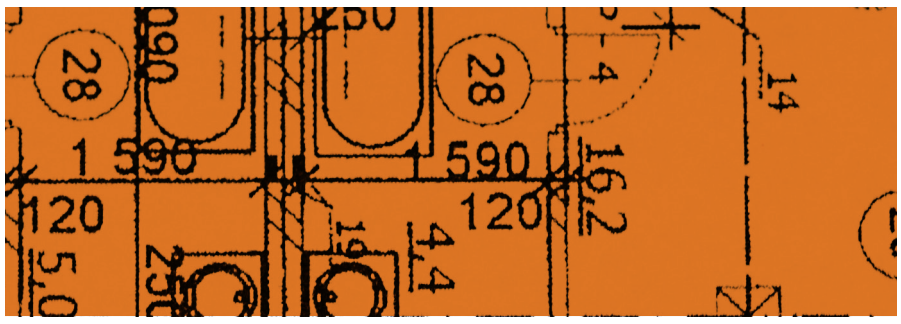
Additionally, as countries seek to modernize trade rules, like-minded nations should band together in coalitions—such as the G-77 or regional blocs like the ASEAN or the Caribbean—to advocate for their interests in larger multilateral forums like the WTO or the Organisation for Economic Co-operation and Development. In a future driven by advancements in AI, data is both air and oil—creating prosperity for everyone.

Camille Stewart Gloster is a global technology and cybersecurity leader specializing in bridging policy and practice to enhance resilience, trust and security. As CEO of CAS Strategies, LLC, she advances a sociotechnical approach that ensures organizations worldwide can build adaptive, defensible ecosystems to navigate the evolving threat landscape and maximize opportunity.

How to build responsible AI for the global majority

AI development needs to prioritize equity and inclusivity rather than deepening existing divides.

Jonathan Julion



As artificial intelligence reshapes industries worldwide, a crucial question is emerging: Who will truly benefit from these advances? While AI offers tremendous potential, there is a pressing need to ensure its deployment is ethical and safe, especially for regions facing unique challenges.

The Global Majority, comprising countries in Africa, Asia, Latin America and the Caribbean, faces both significant obstacles and immense opportunities in the AI landscape. Despite its vast potential, AI faces high barriers in many parts of the world. One of the most pressing challenges is inadequate infrastructure. Countries in the Global Majority often struggle with unreliable internet access, limited data storage and insufficient computational power—factors that hinder the development and deployment of AI systems. Developing AI tools that are accessible in rural or underdeveloped areas is crucial to bridging this gap.

Additionally, there is concern that AI could exacerbate existing inequalities. For example, facial recognition algorithms often fail to accurately recognize non-Western faces due to unrepresentative training data. Similarly, predictive policing systems developed in one country may produce biased results when applied elsewhere. These examples underscore the need for AI systems that are inclusive and designed to avoid perpetuating inequality.

Another critical issue is the lack of comprehensive regulatory frameworks. In many regions, AI technologies are deployed with little oversight, which leads to risks like data privacy violations and algorithmic biases. To ensure AI serves the public good and doesn't deepen existing divides, clear ethical guidelines and governance structures are essential. Despite these challenges, AI offers transformative opportunities. In sectors like health care and agriculture, AI is already making headway in addressing some of the Global Majority's most pressing issues.

In health care, AI has shown great promise in improving diagnostic capabilities in resource-poor settings. In Sub-Saharan Africa, for example, AI is used to detect diseases like malaria, tuberculosis and HIV with greater speed and accuracy than human doctors. These technologies are saving lives and improving health care outcomes, especially in areas with limited access to medical professionals and resources.

In agriculture, AI is helping small-scale farmers in countries like India, Kenya and Brazil optimize their practices. AI systems predict weather patterns, monitor crop health and suggest the best planting times. These technologies are invaluable for farmers who face unpredictable weather and limited access to modern resources, enhancing productivity and reducing risks.

These examples demonstrate that when developed and deployed responsibly, AI can serve as an equalizer, bringing life-saving technologies to underserved communities and transforming traditional industries. However, for AI to reach its full potential, it must be developed transparently, with a focus on the specific needs of the Global Majority.

For AI to succeed in the Global Majority, trust must be at the forefront of its development. Trust is essential for AI's widespread adoption. People must understand how AI systems make decisions, what data they use and how these systems will be applied. Without transparency, AI systems will struggle to gain the trust necessary for acceptance.

To maintain that trust, AI systems must prioritize safety. As AI becomes more integrated into sectors like health care, finance and governance, the risks of biased decision-making, security vulnerabilities and unintended consequences grow. Rigorous testing and clear accountability structures are essential to mitigate any negative impacts.

Creating trustworthy and safe AI is a collective responsibility. Governments, businesses, academia and civil society must collaborate to develop AI systems that serve the public good. This goes beyond technological innovation—it requires new policies, international standards and ethical frameworks that prioritize equity and sustainability.

To ensure that AI benefits the Global Majority, states should implement three policy principles: international collaboration, tailored national AI policies and corporate responsibility. International collaboration may take the form of developed economies investing in AI infrastructure in emerging economies, while sharing knowledge and supporting local innovation ecosystems. To help narrow the AI knowledge gap, developed economies can help build AI research hubs, fund educational initiatives and promote data-sharing.

At the same time, the Global Majority should tailor national AI policies to fit their specific needs and context. These policies should prioritize ethical AI use, data privacy and efforts to reduce inequality. Engaging marginalized communities in policy development will ensure their voices are heard and needs are addressed.

Last, AI companies must prioritize social good over profit. Corporate social responsibility (CSR) should include creating technologies that are inclusive, ethical and sustainable. Companies must commit to auditing their systems for bias and ensuring transparency in algorithm design and deployment.

The future of AI should not be determined solely by the priorities of wealthy nations—it must reflect the diverse needs and aspirations of the Global Majority. By addressing infrastructure challenges, promoting inclusive innovation and establishing ethical frameworks, AI can be a force for good. If we act now, we can ensure that AI benefits people everywhere, fostering a more equitable, trustworthy and prosperous world for all.

Jonathan Julion is an AI ethics expert and philosopher with a concentration in cybersecurity, focusing on the intersection of responsible AI development and digital security.

AI is transforming public services. Here's how to keep it safe and fair.

Automation will cut costs and improve services, but fairness, privacy and accountability must remain a priority.

Mlindi Mashologu

AI adoption in the workplace will reduce many of our most mundane tasks, from document processing to workflow management. It will help cut down on errors and streamline operations, allowing public-sector employees to focus on more complex, value-driven activities, optimizing resource allocation and enabling governments to deliver more with less.

The capacity of AI to analyze data enables governments to provide personalized public services tailored to individual needs, making government services more relevant to individuals and improving the overall citizen experience. Of course, that data should be used ethically, maintaining privacy and trust.

Efficiency should never undermine ethical standards, however. A strong regulatory framework will ensure AI deployment aligns with principles of fairness while sustaining public trust. Transparency is fundamental to trust in AI systems. Strong governance demands disclosure of how algorithms make decisions, enabling citizens to understand and trust these processes. Algorithmic transparency fosters accountability, ensuring biases or errors are identified and addressed. Mechanisms for auditing AI processes ensure that decision-makers and public servants remain responsible for AI-assisted decisions.

Governance frameworks must also ensure that AI systems address the needs of marginalized and vulnerable groups. Inclusive AI initiatives reduce disparities, ensuring equitable access to public services. When citizens trust AI-driven decision-making processes, they are more likely to embrace these technologies, easing implementation and encouraging adoption.

When citizens trust AI-driven decision-making processes, they are more likely to embrace these technologies, easing implementation and encouraging adoption.

The convergence of strong AI regulation and high adoption presents a transformative opportunity for digital government. By prioritizing accountability, inclusivity and public trust, governments can harness AI to reshape public services for the better. The future of AI in governance hinges on balancing innovation with responsibility. Through strategic regulation and ethical deployment, governments can build a digital future that is efficient, equitable and trustworthy, ensuring the benefits of AI are shared by all.

Mlindi Mashologu is a Deputy Director-General: Digital Society and Economy, at the National Department of Communications and Digital Technologies in South Africa, a position he assumed in 2020.

The Internet of Bodies is here. Are we ready for it?

As we implant, ingest and wear smart devices, the line between human and machine blurs—along with our control over our own bodies.

Martina Le Gall Maláková

As uses for artificial intelligence proliferate ever more rapidly within our everyday lives, an ethical question is becoming apparent about the risks and benefits of merging with our technologies.

Like the Internet of Things, which connects objects to the internet, the Internet of Bodies connects you to the internet. The cyborg is more reality than science fiction. We are wearing, ingesting and implanting sensors all over our bodies, from vital health management devices (like pacemakers and insulin pumps) to smart watches. And AI plays a vital role in IoB, mediating the way these technologies interact with our bodies.

Once connected, data from these devices can be exchanged and our bodies could be remotely monitored, controlled and analyzed. The IoB may add value for human life and health, but it also has its risks.

There are, of course, the classic technological risks of cybersecurity and the misuse of our personal data. But there's a secondary risk that's at least as pressing: We risk losing ourselves to the addiction to our devices, and as we merge, what happens when a person is controlled via an AI-powered robot? This could result in loss of control over both one's body and one's self. We must take into account the capacities of use and especially misuse of emerging technologies on society and humanity as a whole.

Frequently, when something new is created, we try to proceed according to established ethical values and principles, adhering to familiar basic rules. But we know from history that in many cases, at a certain point this will not be enough. As Internet of Bodies technology continues to grow, regulatory and legal issues will have to be resolved and policies built around the proper use of the technology.

What happens if IOB devices deviate from their supposed purpose? Or when predatory workplaces start mandating its use to track workers? Who do we hold to account when an AI-controlled device goes awry and causes actual harm?

The European Union's Artificial Intelligence Liability Directive (AILD) is one attempt to answer that question, allowing people recourse when an artificial intelligence system causes damage to them. While the minutiae of the directive is still being debated, opposition to the act is rooted in concern that it will stymie European innovation.

We can already predict that, as we have seen with other emerging technologies, these issues will not be resolved in a timely manner. At the same time, if we limit or slow down the use of these technologies, a gray economy will emerge, where people will be able to purchase unauthorized versions of these tools unbeholden to rules, principles and controls. This is not a solution.

Together, we need to use all possible world organizations, platforms and forums to discuss this topic and, above all, to quickly adopt measures, rules and legislation to mitigate risks. This can only be achieved using a round table, with sufficient awareness of these emerging technologies and sufficient commitment to equality, ethics and human values—for the future survival of our humanity and our planet.

Martina Le Gall Maláková is an expert on Data Free Flow with Trust on behalf of BIAC at OECD.

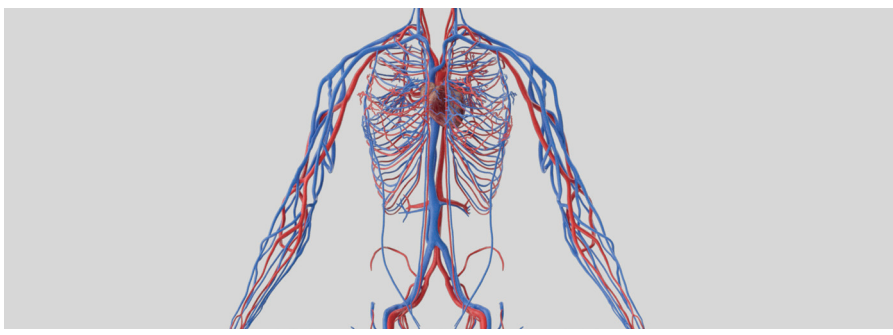


Photo illustration by George Dagerotip

The EU AI act is not necessarily the gold standard for the global majority

The best regulatory approaches will be inclusive and consider regional context and cultural nuances while taking into account how people are most affected by AI.

Paola Galvez-Callirgos

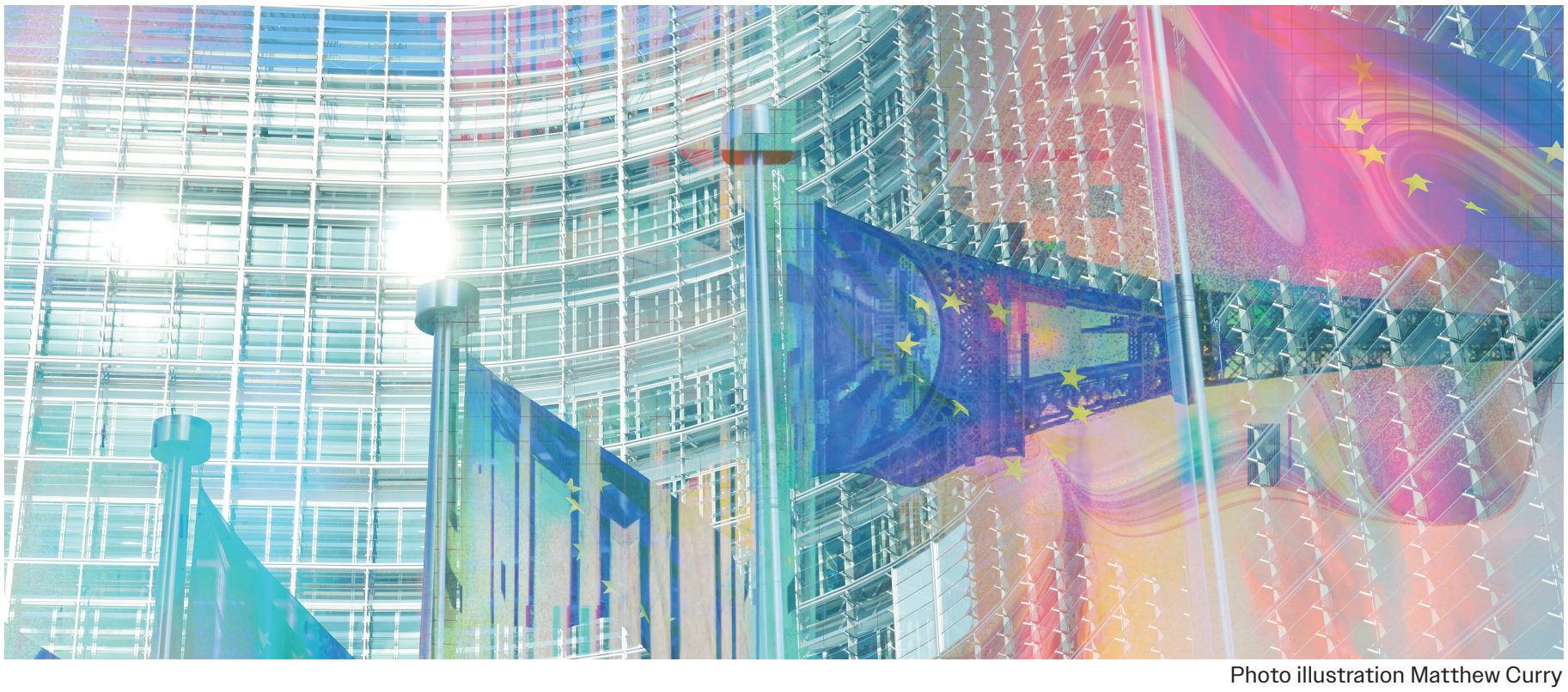


Photo illustration Matthew Curry

As more countries prepare to draft artificial intelligence legislation in 2025, the temptation to merely replicate the European Union model must be resisted. The nuanced challenges of AI governance demand context-specific approaches that reflect each nation's unique technological, economic and social ecosystem, rather than a one-size-fits-all regulatory strategy.

The recent approval of the Global Digital Compact, the upcoming WSIS+20 discussions, and this week's AI Action Summit in Paris have created a pivotal moment for global dialogue. This momentum presents an exceptional opportunity for Global Majority nations to actively participate in shaping AI standards, ensuring that governance frameworks reflect diverse local needs, perspectives and developmental contexts.

To transform AI into an inclusive technological paradigm and put in place the necessary guardrails to make AI benefit all humanity, I propose the following strategic steps to guide policymakers in the Global Majority.

Since each country operates within its unique governance, each should start by building a foundation and integrating globally recognized standards into its national AI regulatory framework. Prioritize adherence to the Council of Europe's AI treaty and the implementation of the UNESCO Recommendation on the Ethics of AI, ensuring ethical, human-centric and transparent development. This alignment fosters international collaboration and strengthens trust and accountability in AI adoption.

Additionally, policymakers should develop or refine the national AI governance framework through an inclusive process that ensures meaningful public participation. AI policies should reflect the perspectives of diverse stakeholders, particularly those most likely to be affected by AI systems. Therefore, after adopting international frameworks, establish a multi-stakeholder bottom-up process and define the nation's unique approach to AI governance. To this end, developing regulatory sandboxes where AI systems can be tested and refined in a safe and supervised manner could be an interesting option to explore.

After this, establish cost-effective independent oversight. Designate or expand the mandate of existing regulatory bodies (e.g. DPA) to oversee AI development and deployment.

Under-resourced nations can build partnerships with international organizations, NGOs and universities to access technical expertise and best practices.

Along with that, build foundational knowledge early by integrating AI and digital literacy into school and university curricula and partner with Development Banks or the private sector to secure funding. Create AI research and innovation hubs in collaboration with universities and business associations to build local expertise and foster entrepreneurship. These actions will build a skilled workforce capable of driving AI innovation and adoption within the local economy.

Guided by UNESCO's recommendations on the ethics of AI, implement targeted initiatives such as scholarships, mentorship programs and capacity-building opportunities for women, Indigenous communities and marginalized groups. Furthermore, frameworks for AI development must explicitly respect local traditions, languages and knowledge systems, leveraging them as sources of innovation. Another factor to consider, mandated by the UNESCO global standard, is to ensure that technological progress does not come at the expense of environmental integrity. Hence, nations shall develop guidelines that promote energy-efficient AI technologies and minimize carbon footprints.

The path to a fair and sustainable technological transition is a complex and extensive journey that requires a holistic approach prioritizing national interests, human capital development and inclusive innovation. These are essential first steps in developing comprehensive and beneficial policy frameworks for AI. I hope this piece is a catalyst for action and provides concrete ideas to start or refine the AI regulatory process in every participating country. It's about time Global Majority countries strategically position themselves as active creators and not merely consumers of AI technologies.

Paola Galvez-Callirgos is a tech policy senior consultant with a master's of public policy from the University of Oxford and researcher at the Center for AI and Digital Policy.

A fair process is *the* essential element for trust & safety

Transparency, consistency, involvement and human oversight are among the critical components of smart content-moderating strategies that don't alienate users.

Louis-Victor de Franssu & Theodoros Evgeniou

Photo Illustration by Matthew Curry

It's more clear than ever that online content moderation is no longer a technical or policy issue—it is a fundamental global governance challenge. Every day, massive digital platforms exercise immense power in determining what speech is permissible, whose voices are amplified and what content is restricted or removed.

Without a commitment to fair processes, this power can be wielded arbitrarily and without scrutiny, ultimately eroding users' trust. And trust is the fragile but essential currency of our rapidly expanding digital ecosystem. The only way to maintain trust is to develop fair content-moderation processes that ensure decisions about removing posts, suspending accounts or any other enforcement actions are made transparently, consistently and with due regard for users' rights. Procedural fairness (how content moderation decisions are made) is just as important to people—and typically more so—than distributive fairness (what decisions are made).

At the moment, content moderation faces a dual challenge: the rapid rise of artificial intelligence-driven systems and mounting regulatory demands. Large language models and generative AI can enable platforms to process vast amounts of data at unprecedented speeds, offering the potential to alleviate some of the key challenges of current automated tools that often struggle with accuracy and context.

While these systems can reduce human biases and improve precision, they are far from flawless, however. They frequently fail to grasp nuance, cultural context and evolving trends in harmful content. As companies rapidly continue to automate, they're bound to face both wrongful removals and undetected violations that are the fault of these systems, underscoring the continued importance of strong human oversight.

At the same time, global regulators are tightening oversight. In the EU, the Digital Services Act requires that platforms implement specific processes and ensure transparency, while granting regulators the authority to monitor compliance. Similarly, the UK Online Safety Act mandates extensive transparency and fair practices, empowering the Office of Communications to hold platforms accountable. Regulation is not limited to Europe. India, Brazil, Singapore and many other regions are adopting new regulatory frameworks. As these regulations take hold, public scrutiny is intensifying, and platforms that fail to uphold fair processes risk both legal penalties and reputational damage.

This dual challenge sheds light on the need to codify fairness directly into the mechanisms of content moderation tools and systems. Fair processes are not just a technical feature or a mere regulatory checkbox. They enable platforms to achieve their core mission and business goals: providing the best experience to users, retaining existing customers and attracting new ones.The keys to developing fair content moderation processes include the following.

Transparency: Users should have easy access to platform policies that clarify why their content was removed or their account was suspended (e.g. the mandatory Statement of Reasons under the Digital Services Act).

Consistency: Moderation decisions should be applied equitably and in a replicable manner across all users, communities and content categories. Both human and AI-based enforcements must be regularly audited with quality controls and live monitoring to ensure impartiality.

Proportionality: The severity of enforcement actions should be proportional to the harm posed by the content, and users should have clarity on the principles underlying such choices. For instance, blanket bans and permanent suspensions should be reserved for the most egregious violations.

Right to appeal: Users should have a meaningful way to challenge moderation decisions, with a structured process for reviewing appeals.

Human oversight: Fair moderation requires human oversight at each stage—design, deployment and evaluation—to facilitate bias mitigation, contextual understanding and adaptability to threats. Real-time monitoring tools should provide access to content and trends overviews, keeping systems in check.

Building fair content-moderation processes is a business imperative that will help platforms attract and retain users by creating the most open yet safe environment for them. Users are more likely to remain engaged on platforms where they feel they are treated fairly, have recourse when moderation decisions affect them and understand how and why decisions about their content are made.The companies that prioritize fairness, transparency and accountability will not only navigate the evolving technological and regulatory landscape more effectively but will emerge as industry leaders.

Louis-Victor de Franssu is the co-founder of Tremau, an online trust and safety company, and the former deputy to the French Ambassador for Digital Affairs.

Theodoros Evgeniou is a professor of technology and business at INSEAD, where he directs the executive training programs on AI, and is a co-founder of Tremau

‘AI for Good’ shouldn’t become the new innovation arbitrage

If advanced technology is being tested more aggressively in the world’s most under-resourced communities, it’s worth asking why.

Rumman Chowdhury

At the AI Action Summit in Paris this week, tech leaders and civil society alike will share stages to tout the benefits that artificial intelligence can bring to the Global Majority. When asked for specifics on how AI will improve humanity, CEOs and tech leaders overwhelmingly point to improvements in industries like health care and education in low-resource regions. However, without careful attention to how this tech is deployed, we run the risk of using the Global Majority as a testing ground for incomplete AI solutions that exacerbate disparities in access and quality of care in these industries and others.

It’s understandable to see why AI in education and health care may be appealing. It is difficult to find qualified individuals willing and able to live and work in impoverished regions, and locals who do receive medical or educational training often seek more economically secure and upwardly mobile employment prospects elsewhere. Those that stay often end up being overworked. Inevitably, a teacher handling a classroom of 50 children may not have the resources to work with a child with a learning disability. Medical professionals may not be equipped to diagnose complex medical issues with more obscure symptoms.

AI is well poised to close these gaps. We have seen prototypes of customized learning AI tutors providing unlimited educational content. We have seen telemedicine enable rural doctors to perform surgeries and improve patient care. However, if these technologies are predominantly used in lower-income neighborhoods but applied more carefully in affluent ones, it warrants extra scrutiny. Deployed poorly, so-called “AI for Good”—or the deployment of automated systems meant to have a social impact—can serve as a convenient test environment for these tools in a less visible or legally restrictive environment, with local populations as collateral damage.

The term “innovation arbitrage” refers to the careful exploitation of regulatory gaps to deploy questionable technologies. More resourced economies may have stricter regulations around labor, privacy or data management. It becomes easier to test novel technologies in less developed economies with the added benefit that it’s unlikely that egregious failures in rural regions will be picked up by major media the way they would for a more affluent population. Uber is a perfect example. The company had an explicit strategy to test manipulative algorithms to improve driver productivity in countries where worker protection laws were weakest, while paying as little as possible.

Any concept of equity requires an assessment of quality. Expecting the impoverished to be grateful for any handout, delivered enthusiastically but executed poorly, is a reflection of the savior complex that often pervades tech-for-good communities.

Innovation arbitrage allows companies to gather data to improve their algorithms with little to no consequence, and enables them to refine their product for more affluent markets, where data protection laws, liability laws and more can prevent them from at-scale testing. Without the appropriate safeguards, AI for Good initiatives can be easily manipulated to serve as a new method of innovation arbitrage.

The problem is twofold. First, our definition of “digital public infrastructure” focuses on “capacity building”—that is, data collection and curation, computational resources and access to models. We rarely discuss tools for testing and evaluation to determine if these products are safe, secure and responsibly deployed.

AI for Good initiatives offer large prizes of computational resources and funding for technology teams who are passionate about the problem they’re working on and are embedded within their communities. However, these teams often skip hiring individuals who are best able to test for security, privacy and ethical flaws. It is assumed that technology teams will fix such issues post-deployment if they deem them critical enough. This means that whether these mistakes are prioritized is not driven by any legal or ethical imperatives to protect consumers, but by staffing capacity and funding. At the same time, the companies that produce these foundational models are able to gather hard-to-acquire test data from these communities.

Second, generative AI algorithms introduce a host of possible failures, far beyond the limited algorithms deployed by Uber and others in earlier iterations of innovation arbitrage. These AI tools can confidently reflect embedded biases or completely hallucinate responses. Without adequate quality checks or clearly defined methods of testing, developers risk solving the most visible problems instead of the most impactful ones.

My nonprofit, Humane Intelligence, has conducted tests to examine biases within AI tools with a wide range of communities, and has identified significant issues that can arise when well-intended technologies are launched with inadequate testing and evaluation. Testing methods for Generative AI models are poorly defined, with no clear standards or thresholds for harm identification or mitigation. The few tests that exist are largely Western-focused, for example, identifying “bias” by American racial constructs while ignoring forms of non-Western bias like caste. These limitations have real-world implications.

Along with Singapore’s Infocomm Media Development Authority, we examined a range of large language models that demonstrated significant regional biases in an Asian context. These biases arise because the individuals training and testing the core models on which educational bots are built simply lack local cultural context. Imagine an AI tutor deployed in rural parts of Malaysia that confidently states that individuals from the primarily rural eastern regions are less hygienic and more likely to make poor economic decisions because of their cultural shortcomings.

Photo Illustration by Allison Saeng

Utilizing AI as a tool of equity requires more than determining if a problem is answerable at scale. Funding data, computational resources and model development are necessary starting points, but equitable care does not mean simply providing access where there was none before. Any concept of equity requires an assessment of quality. Expecting the impoverished to be grateful for any handout, delivered enthusiastically but executed poorly, is a reflection of the savior complex that often pervades tech-for-good communities. In addition, enabling unfettered AI deployment opens the door to a global testbed of subjects to be manipulated at will, in a revived form of innovation arbitrage. This is not to say AI cannot be used for good, but that AI for Good investments also need to fund context-specific evaluation, safety and security measures.

Dr. Rumman Chowdhury is CEO and co-founder of Humane Intelligence, and the first person to be appointed by the Department of State as the United States Science Envoy for Artificial Intelligence.

Smarter innovation in a box

Thiago Guimaraes Moraes

Safe and responsible AI development requires careful testing and regulation. That's why AI regulatory sandboxes—controlled environments for experimenting with innovative technologies and data practices—are a promising tool for the Global Majority to allow for technology leapfrogging while fostering responsible innovation and ensuring that new systems respect fundamental rights.

These sandboxes serve a dual purpose: They are secure environments for testing technologies, and they are collaborative frameworks that allow regulators, innovators and other stakeholders to assess technologies against existing laws and ethical standards. Sandboxes are time-limited, iterative processes designed to evaluate innovation while ensuring regulatory alignment.

The benefits of sandboxes extend beyond innovation. They provide a structured mechanism to balance the trade-offs between fostering technological advancement and maintaining regulatory compliance. By encouraging the pooling of datasets and expertise, sandboxes allow participants to explore new possibilities while addressing concerns around privacy, fairness and accountability.

For the Global Majority, regulatory sandboxes offer a strategic approach to bridge the gap between innovation and compliance. By fostering localized benefits and demonstrating the tangible value of AI investments, sandboxes can attract international partnerships and funding. Regional pilots, enabled by these environments, showcase the potential of AI to address local challenges, from optimizing public service delivery and access to advancing healthcare solutions tailored to specific contexts.

For example, in Africa, the nonprofit Datasphere Initiative has organized sandbox discussions to build a pan-African community

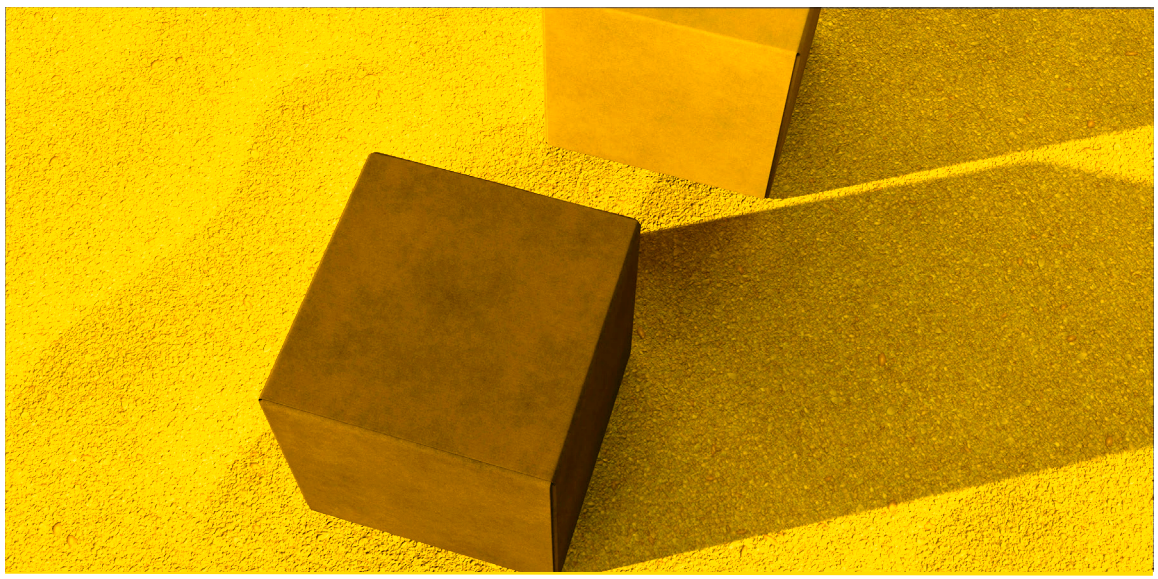


Photo illustration by Allison Saeng

to enable innovative cross-border data governance solutions. The first iterations focused on operational sandboxes, and as the initiative moved forward, it facilitated discussions to create regulatory capacity and connections via roundtables.

Implementing sandboxes in the Global Majority is not without challenges. Limited resources often hinder the establishment of effective initiatives. Without proper planning and capacity-building, these efforts risk becoming time-consuming exercises with little regulatory learning or meaningful outcomes. Another challenge is ensuring that sandboxes are not just experiments but effective platforms for real innovation. Poorly designed sandboxes can lead to outcomes that neither advance regulatory understanding nor deliver societal benefits.

To maximize the benefits of regulatory sandboxes, the Global Majority must adopt a strategic approach that leverages partnerships and builds on existing strengths. Consortiums involving governments, private sector players and international organizations can pool resources and expertise, addressing resource constraints of developing economies. In an increasingly interconnected world, the Global Majority has a unique opportunity to lead by example, demonstrating how AI can be harnessed to address local challenges while adhering to global principles.

Thiago Moraes is a joint-degree Ph.D. Candidate in law at University of Brasilia and Vrije Universiteit Brussels and a PhD fellow of the Digital Governance cluster at the United Nations University's Institute on Comparative Regional Integrated Studies.

Who will save us from a future without work?

With AI expected to alter or eliminate nearly 40% of global jobs, the risks of mass unemployment and economic disruption loom large. But there's a better way to reshape our economy to benefit workers instead.

Alexandra Samuel

Consider a world in which work is endlessly meaningful and creative, free of rote drudgery or backbreaking labor. At last year's TED conference, Daniela Rus, director of MIT's Computer Science and Artificial Intelligence Laboratory, held up this possibility. "When AI moves into the physical world, the opportunities for benefits and for breakthroughs [are] extraordinary," Rus promised, underlining her vision with images of robots carrying groceries and delivering packages.

We may have a few years before robotics and AI bring that particular question to our collective doorstep, but AI is already reshaping work. The IMF has estimated that AI may eliminate or change 40% of the global workforce and as much as 60% in advanced economies. That doesn't mean that 40% of jobs are going to disappear, but it does mean we are in for an extended period of turbulence and transition that will affect a great many people.

Amidst all the excitement (and hand wringing) about AI's long-term possibilities, I wanted to know how AI leaders think about these more certain and near-term risks: the elimination of many jobs, the pain of economic restructuring and the possible rise in overall unemployment. How worried should we be, and how soon?

In his book "A World Without Work," economist Daniel Susskind points out that a dramatic rise in German unemployment (to 24%) was part of what brought Hitler to power. I put the question of whether AI's impact on the economy may lead to civil unrest to Tom Gruber, an attendee at the Vancouver TED conference who is also the co-founder of the company that created Siri, before it was acquired by Apple. Gruber, who now advocates for "humanistic AI" as a speaker and impact adviser, isn't too concerned about the current wave of generative AI chatbots displacing high-skilled employees—yet.

"We can talk to these [chat]bots, but we should not be trusting their wisdom," he says. "They're like a 22-year-old fresh out of college telling you an opinion they've acquired after three years of drinking and talking. They're just not going to really solve business problems with expertise."

Even with those limitations, Gruber notes, there's lots that bots can already do better than humans. He gave the example of high-end marketing work: While humans are still much better at ideation, when it comes to generating marketing assets like mockups, the AIs "totally kick butt on the humans." The sheer volume of work these chatbots can generate at high speed, he notes, "is going to put downward pressure on wages."

Under our current labor conditions it's hard for employees to fight that kind of pressure. "There's the ubiquity of AI everywhere, always, listening to everything always," Nita Farahany, a Duke professor and author of "The Battle for Your Brain," tells me after the conference. The sheer volume of data that AI consumes, according to Farahany, means it's only a matter of time until it becomes capable of replacing more people—and we're embracing AI at a pace that leaves little room for addressing that human impact.

There is still time to plan for that kind of workforce re-skilling and reorganization, however, so that displaced workers aren't simply dropped from the workforce. Governments might play a role in creating room for a more careful transition: California's state legislature recently passed a bill that would have prohibited the government from outsourcing work to call centers that use AI to replace human workers, but Governor Gavin

AI safety is a misnomer without Global Majority inclusion

A limited understanding of AI's real-world impact on the Global Majority means the world's most populous and under-resourced countries remain at risk.

Chinasa T. Okolo

Artificial intelligence safety has emerged as a critical area of inquiry, seeking to ensure that systems operate reliably, ethically and beneficially. However, mainstream AI safety discourse remains largely shaped by Western objectives and priorities, often privileging concerns such as technological alignment and misuse over broader societal and contextual harms.

This narrow framing is reflected in broader concerns about the risks of AI and in the focus of AI fairness and safety research, much of which is produced within and primarily addresses the contexts of Western, Educated, Industrialized, Rich and Democratic (WEIRD) countries. As a result, AI safety efforts are disproportionately designed to benefit stakeholders in high-income nations, frequently neglecting not only the lived realities of marginalized populations within these societies—including Black, Latinx and Indigenous communities—but also those in the Global Majority.

This exclusion is not a small oversight; it risks deepening existing global inequities by failing to account for how AI systems process and interpret non-Western languages, cultures and values while simultaneously amplifying risks for Global Majority communities.

Recent international initiatives, notably the series of AI Safety Summits, have sought to address global AI safety concerns. However, these efforts continue to demonstrate limited inclusion of Global Majority perspectives. In 2023, the UK hosted the inaugural AI Safety Summit to convene government officials, representatives from top AI companies, civil society stakeholders and academic researchers to discuss the risks of AI and to work toward mitigation through technological and regulatory measures. Yet out of the 27 governments represented, only seven—Brazil, India, Indonesia, Kenya, Nigeria, the Philippines and Rwanda—were from low- or middle-income countries.

The AI Seoul Summit in May 2024 saw even less representation, with only three Global Majority countries (India, the Philippines and Rwanda) among the 20 governments in attendance. While it has been unclear which countries will actually participate in the Paris AI Action Summit, a significant increase in Global Majority participation remains uncertain. The launch of specialized AI Safety Institutes further exemplifies the lack of Global Majority participation: While Kenya is represented, the majority of network institutes are based in high-income nations, including the UK, US, Japan, France, Germany, Italy, Singapore, South Korea, Australia, Canada and the European Union.

A significant dimension of AI safety has centered on the advancement of technology benchmarks, which serve as crucial tools for evaluating the capabilities and associated risks of AI systems. However, many of these benchmarks are underpinned by Western-centric assumptions regarding trust, safety and security, which are subsequently reinforced in evaluation methodologies.

Widely used benchmarks like the Massive Multitask Language Understanding exhibit limited coverage of non-Western languages, topics and cultural norms. This gap, compounded by the failure of general-purpose AI developers to enhance the cultural robustness of their systems, results in AI models that perform inadequately in diverse contexts and therefore perpetuate systemic biases. Moreover, technological solutions such as automated content moderation are primarily optimized for English, leaving other linguistic communities vulnerable to misinformation, censorship and harm. These deficiencies underscore the urgent need for AI safety frameworks that account for the linguistic, social and political complexities of the Global Majority to enable the development of safe and reliable AI systems.

The limited empirical understanding of the real-world impact of AI in Global Majority contexts further constrains efforts to develop inclusive safety strategies. To move toward more inclusive AI safety, we must first work to understand the risks AI systems pose to populations and consumers within the Global Majority.

A significant amount of AI fairness research focuses on Western contexts and revolves around Western constructs such as race. While this has yielded important insights into facial recognition bias, it remains insufficient for capturing the multifaceted nature of AI-related harms in non-Western societies. Factors such as caste, tribal affiliation, religious identity and their intersection with other dimensions of social stratification, including gender and socioeconomic status, play a crucial role in shaping the lived experiences of communities in the Global Majority. As AI systems continue to be deployed at scale in these regions, it is imperative that the international community—including frontier AI developers, international standards bodies and multilateral institutions—prioritize a more holistic and contextually grounded approach to AI safety.

Advancing an inclusive AI safety paradigm requires meaningful investment in Global Majority-led research and capacity-building initiatives. Researchers from these regions must be provided with adequate resources to develop contextually appropriate evaluation methodologies. Additionally, Global Majority governments must be equitably represented in international AI governance discussions and afforded substantive opportunities to shape the trajectory of AI safety initiatives. Globalized approaches to AI safety provide a critical opportunity to reshape discourse and practices around responsible AI. By centering the unique risks, opportunities and cultural considerations of Global Majority communities, these efforts can redefine what it means for AI to be "safe" in a pluralistic society. Addressing longstanding structural imbalances in AI safety discourse, alongside intentional investments in research and advocacy for globalized AI safety approaches, will require sustained commitment but is essential to fostering a more equitable AI future.

Chinasa T. Okolo, Ph.D., is a fellow at The Brookings Institution and a recent computer science Ph.D. graduate from Cornell University. Her research focuses on AI governance for the Global Majority, datafication and algorithmic marginalization, and the socioeconomic impact of data work.

Newsom vetoed the bill.

Another possibility is for employers themselves to take responsibility for finding new jobs or tasks for employees whose work is automated. When IKEA announced it would use AI to take over the work of its call center operators, for example, it retrained the displaced operators as interior design advisors.

That's the kind of approach that's championed by Chet Kapoor, the CEO of DataStax, a database company that powers AI applications. Earlier this year DataStax published a white paper on how AI could turn into a win-win for both employees and employers. Kapoor argues that rather than using AI to lay off workers, smart employers may even increase headcount as AI makes each worker more productive.

Take the case of programming talent. AI has already proven so effective at coding that some industry leaders—like Matt Garman, the CEO of Amazon Web Services—are predicting that AI will take over all the work of actually writing code. But Kapoor says that at least for the next decade, employers have more to gain by expanding their coding teams to take advantage of generative AI.

"It doesn't matter whether it's a tech company or a non-tech company," he says. "There's not a single company that doesn't have a backlog of apps that they want to get done. Let's go and build those apps."

Kapoor acknowledges that there are some companies that will use AI to cut headcount and costs. But he argues that there are also employers who recognize generative AI as an opportunity to accelerate or innovate. If AI leads to some job displacement, employers can work with programmers to reskill and redeploy talent. In this version of the future, the expanding opportunities of AI solve the problem of job displacement. As AI increases productivity, we can do more, make more and sell more, without any need to shrink the workforce at all.

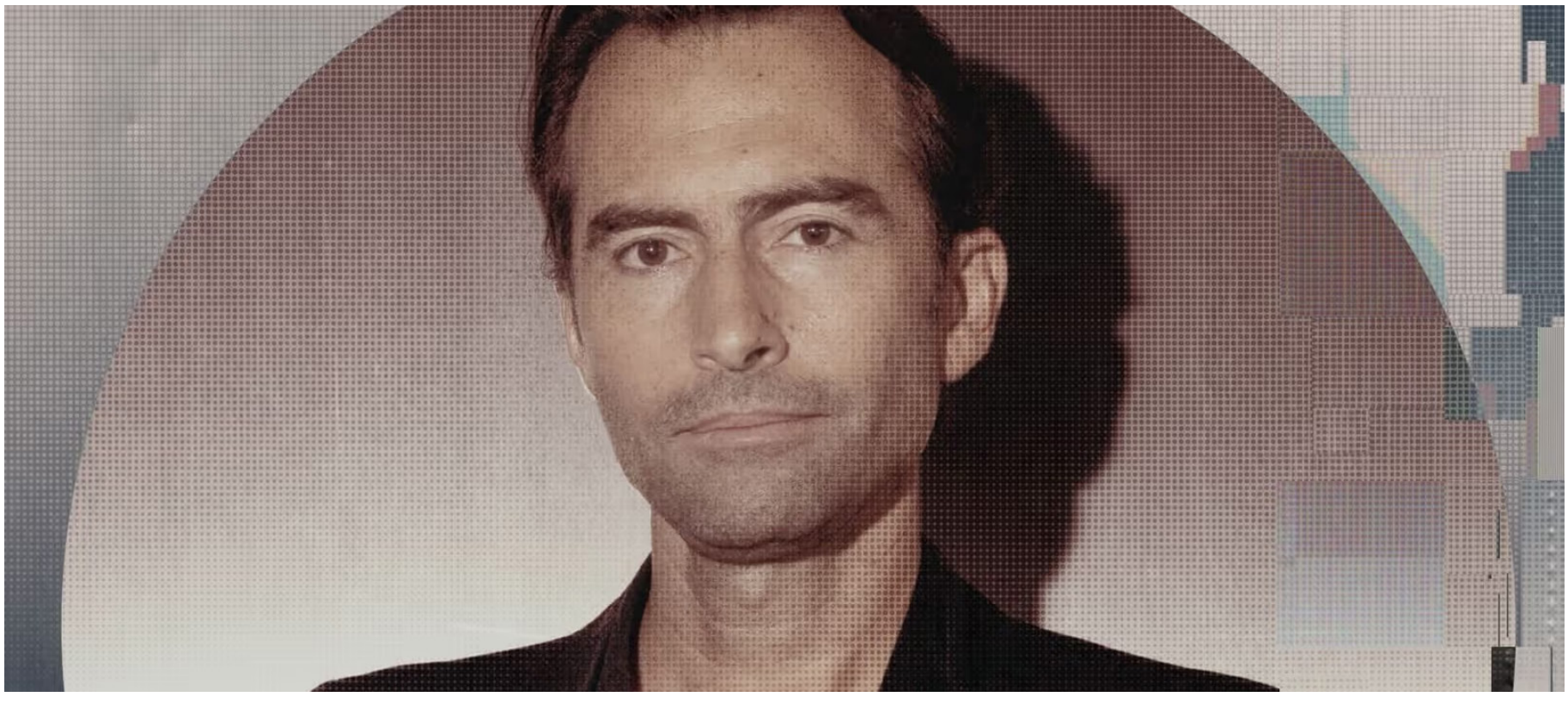
What happens if markets can't scale that fast? Consider a second option: Use the productivity gains from AI to reduce working hours, but without reducing compensation. If we pay people based on output rather than hours, then AI-enabled efficiency could make it feasible to keep salaries constant while reducing hours spent working. In this scenario, AI would effectively increase hourly wages.

It may sound like a radical idea, but it's one that has already proven successful. In her TED talk on "good jobs," MIT professor Zeynep Ton pointed to the success of the bulk retailer Sam's Club, which boosted productivity, reduced turnover and drove membership growth, all by increasing hourly pay. Providing people predictable, manageable hours—as opposed to burnout-level schedules or second jobs—is what makes work sustainable and satisfying.

The artist fighting for humanity in the AI era

John Mack, a tech critic and visual artist, warns against our growing reliance on artificial intelligence and digital tools and makes the case for a more human-centered future.

Nancy Scola



Visual artist John Mack at the opening of his show “A Species Between Worlds” on Sept. 8, 2022, at the Skylight Modern Gallery in New York City. Photo by Jason Sean Weiss/BFA.com. Photo illustration by Matthew Curry.

“We have a parasite,” warns John Mack on a rainy Thursday afternoon this past spring. We’re on the second floor of the Explorers Club in Manhattan’s Upper East Side, where Mack is giving a talk to a few dozen guests. This is a place more used to ruminations on the ravages of trichinellosis than the bloodsuckers Mack is here to discuss: modern technology, of the likes of Facebook, TikTok and X.

Founded 120 years ago by arctic explorers, the mahogany-paneled club plays host to those who’ve journeyed the world and come back with lessons to share. Mack is a street photographer turned immersive visual artist turned interdisciplinary tech skeptic, and is a 2022 member of the Explorers Club’s 50 (“fifty people changing the world that the world needs to know about”). He’s exhibited at the London Design Biennale and serves on the board of the nonprofit Fairplay, which seeks to protect kids from the ills of Big Tech. And he gravitates toward big ideas and creative ways of explaining them.

Twenty-first-century humanity, Mack declares, has formed a mutualistic relationship with our digital tools: It provides entertainment when we’re bored and fulfillment when we feel empty, and we provide it with a business model. “The stronger the parasite, the stronger the host. The stronger the host, the stronger the parasite,” he says. “That’s what we’re dealing with here.”

But unlike the luddites of yesteryear, who were more concerned with advancing technology’s impact on our livelihoods, Mack’s worries are more fundamental. The challenge, he suggests, is that we haven’t stopped to wonder whether that interdependency between humanity and our technology is making us at once happier and yet less human.

Mack had initially, through a representative, agreed to an interview, but he ducked my attempt to chat after the presentation. During the Q&A portion of the session, however, he seemed open to entertaining questions from the audience. A person named Debbie, who says she was left “slack-jawed” by Mack’s presentation, asks whether he believes Big Tech’s big plan is to effectively turn us into cyborgs, an amalgam of human bodies and human-made machinery.

Mack, tall, goateed and in his late 40s, wrestles with the question. He flirts with seemingly far-fetched ideas, at one point suggesting that tech companies might want to conscript us into some unspecified wars they might soon be itching to fight. But he lands on the notion that whatever their motivation, the logical outcomes of their actions should be enough to give us pause. “I do think that humanity is going to split,” he says. “It will be those who will stay human, and those who will become more machine-like. I mean, it really is ‘Star Wars.’”

As Mack sees it, it’s almost too late. As he put it in a talk at Stanford’s Institute for Human-Centered Artificial Intelligence last year, “We’re approaching a tipping point where, once crossed, we’ll never be able to return to what it means to be human.” The key, then, he argues, is getting people to see—now—what’s at stake.

Mack is hardly alone in this effort to wrestle with one of the most pressing questions of our times: What exactly is technology doing to our humanity? But to an exceptional degree, he has thrown himself into pursuing a satisfying answer.

Mack, originally from New York City, graduated from Duke in the late ‘90s and jumped into a career in the visual arts. He served as a production assistant on a documentary film shot in Antarctica, and produced art books on Mexico and Marseille. He currently splits his time between London and Sevilla, Spain.

But eight years ago, in 2016, he found himself alarmed when his eye caught on a bit of news set in the streets of the Taiwanese capital city of Taipei. A veritable stampede of people rushed after a rare character, called a Snorlax, in the hybrid virtual-offline “Pokémon GO”’ game. It looked, thought Mack, like a mass migration, a natural flow of life pulled into an artificial space, leaving humans as refugees caught between two worlds.

Mack set out to capture what, exactly, that looked like. It took more than 200 airplane flights, seven helicopter trips and one husky-drawn sled, but he traveled the world, hitting 50 U.S. national parks and each of the world’s “Seven Natural Wonders,” from Victoria Falls to the Great Barrier Reef. During his travels, he captured pairs of images: what the iconic settings looked like in a traditional landscape photograph, and what they looked like in digital rendering through the “Pokémon GO” app. The images formed the foundation of an immersive exhibit—a “gamified meditation,” Mack described it—called “A Species Between Worlds: Our Nature, Our Screens.” TimeOut New

York called it a must-see.

But Mack wanted to do more, to offer his fellow human beings a way of navigating the confusingly eroding boundaries between real life and the digital realm. In 2021 he founded a nonprofit, The Life Calling Initiative, aimed at the mission of helping to "preserve our humanity."

It's no easy task when there are today more mobile phones in the world than souls. In a slim book of notes and aphorisms called "Notes to Selfie: Bits of Truth in a Phoney Word," Mack quipped "I, phone."

In trying to meet this challenge, Mack's put himself in good company. There is, for example, President John F. Kennedy, who in 1963 resisted the idea that growing workplace automation was necessarily dehumanizing by putting the onus on humans to stick up for themselves. "I think machines can make life easier for men," Kennedy said, "if men do not let the machines dominate them."

More recently, Virginia Wesleyan professor Steven M. Emmanuel has drawn lessons from the career of PBS legend Mister Rogers, concluding that the key is to fight back against communication technologies' attempts to make our dealings with other humans crasser than they are offline. We need to return to a time, argues Emmanuel, of "owning your own words."

Then there's the billionaire businessman and onetime owner of the Los Angeles Dodgers, Frank H. McCourt Jr., who argues in his new book "Our Biggest Fight: Reclaiming Liberty, Humanity, and Dignity in the Digital Age" that tech companies hoovering up our personal data isn't about invasion of privacy, it's about stealing our personhood. "Digital feudalism," McCourt calls it, calling for the creation of an alternative internet with respect for individuals' data autonomy baked in.

As Mack sees it, the world is waking up to the fact that we haven't gotten the technology-human balance quite right, and that imbalance has left us all a little off. He notes that there is "a lot of unnerving energy frequency right now around the planet," which is connected to the destabilization of democracies all over. "Somewhere in the depths of souls, we know this is happening," he insists.

That said, Mack isn't about to call for abandoning technology. Rather, his approach alludes to Luddism, in its truest sense. As former LA Times columnist Brian Merchant, author of the 2023 book "Blood in the Machine: The Origins of the Rebellion against Big Tech," has put it, the 19th-century Luddites "were not, contrary to popular belief, idiots who broke machines because they didn't understand them." They simply wanted to harness machines to their benefit.

"I am not anti-tech," Mack said during his Stanford talk. "I am pro-humanity." In fact, he insists, he's excited about artificial intelligence, including because it will help us sort out what's truly the domain of machines and what abilities belong solely to humans.

Designing technology with the well-being of humans top of mind has of late gained major traction in tech circles; one of the leading organizations in the space is called the Center for Humane Technology. In the past few years, lawmakers have suggested imposing regulatory correctives on our machines, from banning micro-targeted advertising to reviving US antitrust laws to make sure AI makers compete on human-respecting products.

But while such moves, Mack argues, might be useful "repellants," they aren't cures for what ails us. So he's landed on a different fix.

The answer, suggests Mack, is strengthening humanity to be better able to contend with technological parasites. The Life Calling Initiative's new project is something Mack calls h+1—as he puts it, a "humanity impact fund." It came about, he says, when speaking with a friend in California who wanted to put $10 million into a fund aimed at teaching empathy and compassion to artificial intelligence.

"I said, 'That's great,'" recalls Mack. "But can we at least take half that, and put it into a fund that teaches humanity empathy and compassion?"

The "h," then, in h+1 stands for humanity. The +1 represents the idea of developing a technology-educating fund that goes to organizations doing something to foster human-centered connection. He's on the hunt for organizations that nurture "lines of connection to the human soul," an echo of 19th-century Romantics' spirit-strengthening response to the Industrial Revolution. The fields in which such organizations work don't matter—it could be philanthropy or education or health care, says Life Calling—as much as that they are working on deepening connections in one of four areas: to self, to others, to nature,and to imagination.

The future is now

There's no time to waste, Mack warns, including because the pace of technological development is at the moment so fast that time is effectively shrinking. "Look at how quickly AI is moving," he says. "What was [once] way out in the distance right now is exponentially getting closer to us … . The time thing," he says, "is an issue."

Mack draws our attention to cards left on our chairs with a QR code that we can scan to open a form through which to nominate a worthy h+1 recipient. A short time later, Mack wraps the session and we humans—passing by the hulking remains of Percy, a stuffed adolescent polar bear—repair to the library for conversation and light refreshments. Mack's call to action has only grown, especially as, he says, technology seems to be winning the battle for attention.

In June, during a presentation in Amsterdam called "AI and the Unconscious Migration: Investing in the Inner" hosted by the Mozilla Foundation, Mack pointed the crowd's attention to a Bloomberg News piece: "Kamala Harris set to unveil $200 million AI investment from private foundations."

"Instead of just seeing this news announcement," Mack said during his Amsterdam talk, "wouldn't it be wonderful to see a headline like this: 'Philanthropy unites on $200 million humanity fund to hedge for AI future?'"

Nancy Scola is a magazine writer based in Washington D.C. whose work often focuses on the intersections of technology and politics. Her deeply reported feature stories on the role technology plays in modern politics and the history that informs it have appeared in publications like New York, Wired, The Information, Washingtonian and The Atlantic. She is a contributing writer at POLITICO Magazine.

The Trump administration's war on facts is now underway. Here's how to fight back.

A global alliance is essential to countering disinformation that prevailed during the campaign and will continue to pollute the internet thanks to Trump allies Elon Musk and Mark Zuckerberg.

Michelle Daniel

Photo Illustrations by Matthew Curry

As President Donald Trump begins his second term, pundits, analysts and democratic strategists will put forth a string of theories about what led to his return to the White House. Central to their discussions is the role of "new media"—podcasts, niche online news sites and social platforms—that has far surpassed legacy media in reach and influence, and Trump's seemingly effortless domination of this vast, chaotic ecosystem. It certainly helped him secure the election.

But the real winner this past November wasn't Trump. It was disinformation.

Disseminated and amplified by the very platforms that were designed to foster connection, disinformation has emerged as one of the most insidious threats to our democracy. Unlike misinformation, which is information that's inaccurate accidentally and without ill intent,

disinformation is deliberately untrue and misleading. President Joe Biden in his farewell address emphasized this concern saying "Americans are being buried under an avalanche of misinformation and disinformation, enabling the abuse of power."

To seal the death of facts, one of Trump's first executive orders mandated the end of what the administration described as the government's censorship of social media—a manufactured crisis that the far-right has used to target disinformation researchers and social media platforms through lawsuits and congressional inquiries over the last several years. "Under the guise of combatting 'misinformation,' 'disinformation,' and 'malinformation,' the Federal Government infringed on the constitutionally protected speech rights of American citizens across the United States," the order reads. Trump also promised to investigate government agencies engaged in these activities, putting a stark end to federally funded research into mis- and disinformation.

The instability of our democracy, exacerbated by unchecked recklessness with information and a woefully inadequate education system, is glaring. Disinformation causes decay indiscriminately. No telling what America will look like in the next election cycle if we allow disinformation to negate truth, pollute online environments and erode public trust in institutions. To understand what's at stake, let's explore the current checks that will likely disappear under the new administration. Then we'll discuss how we might salvage the remnants to stay in the fight.

One of the cornerstones of the Biden administration's fight against disinformation was the State Department's Global Engagement Center. Founded in 2011, the GEC fell under the jurisdiction of the Under Secretary for Public Diplomacy. The agency's original mission was to counter false narratives about the United States and harness social media data to understand how foreign disinformation campaigns impact perceptions of the United States abroad. Lee Satterfield currently holds the position of Acting Under Secretary for Public Diplomacy, but she's unlikely to be confirmed before the end of the current administration. Even if she were, the GEC closed permanently on Dec. 23, 2024 after being perceived by Republicans as a leftist censorship machine.

In February 2023, Elon Musk provided a glaring example of the effort to discredit the GEC, which now stands out as most germane to the incoming administration. Via Twitter/X, he labeled the GEC as not only a threat to democracy but also "the worst offender in U.S. government censorship and media manipulation." Tellingly, this accusation came just as the GEC decided to escalate its counter-disinformation efforts targeting Kremlin-funded campaigns. Another episode around this time involved Republican members of Congress—Michael McCaul (TX-10), Brian Mast (FL-21) and Darrell Issa (CA-48), among others—who sent an oversight letter to Secretary of State Antony Blinken arguing that the GEC had violated its original counterterrorism mandate and had been censoring disfavored viewpoints particularly from conservative American media.

The GEC had created a framework for countering malignant foreign information in the hopes of building consensus among nations for this universal problem. It signed memoranda of understanding with other countries, attempted to build bridges with like-minded counterparts in academia and sought to create interagency collaboration for the purposes of disarming disinformation and propaganda. While the GEC might have been encumbered by bureaucracy and politics that limited its efficacy and action, it did in fact have an impact. And though a replacement body is in the works, this new office will be strictly focused on foreign interference and information manipulation. This precludes all the homegrown destructive narratives throttling the domestic digital space that creates impediments for local, state, and federal authorities simply seeking to do their jobs—as we have now seen countless times during climate disasters from Los Angeles to North Carolina.

Protecting our information environment goes hand-in-hand with media literacy. Any micro-momentum that has been made by the Department of Education toward getting school districts nationwide to integrate media literacy and cybersecurity training into the standard curriculum is going to be undone or halted. Education will overall be in crisis mode, and any chance we have of systemic change to keep pace with the needs of our tech-dependent workforce will fade—at least for a time.

Some of the biggest traditional problems with tackling misinformation and disinformation are the fragmented, whack-a-mole approaches to "hostile information activities," as NATO calls them. Working in silos does not lead to long-term change. The military does not send soldiers into kinetic warfare one at a time and separately without shared intelligence and operational support; tackling disinformation should be no different. The only way to begin to "win" the information war is through a cross-denominational, future-focused and horizontally structured collaboration in the form of an information working group (IWG). Disinformation researchers can take a cue from the Climate Action Network, a network of nearly 2,000 civil society groups in over 130 countries that collectively fight for sustainable solutions to the climate crisis.

Why future-focused? A compelling vision of the future is a uniting force. Why horizontal? Cross-cutting networks are a deterrent to any one entity's domination, and they connect many universities and researchers, cross-pollinating them with NGOs, advocacy groups, industry and intergovernmental organizations.

To guide the labor of this IWG, there must be a new kind of grand strategy. I dusted off my grad-school copy of "The Chessboard & the Web," wherein Anne-Marie Slaughter proposes a pleasant grand strategy of Open Order Building. But given the extent to which the online ecosystem is independent of geographical boundaries, we shouldn't be focusing on a state-centric approach of any kind. We should look away from government-funded institutions that are bound by geographical constraints: Borderless information requires borderless leadership and borderless strategy. Information disorder is universal. We need to move on from George Kennan's "containment" or Biden's "constrainment" and the overt democracy promotion inherent to each. Democracy is only as strong as its people. We should instead be guided by an urgent need to foster networks of collaborations while strengthening global media literacy programs.

A significant impediment to the IWG model, which will certainly discourage essential cross-border collaboration, is the legislation allowing the Treasury Department to shut down 501(c)(3) nonprofits, as many key players rely on their nonprofit status to operate and fund disinformation research. To ensure resilience, the IWG must include international partners and diverse funding mechanisms beyond the U.S. government that can survive domestic political vicissitudes. Only a network of networks rooted in trust and collaboration can hope to counteract the destabilizing power of disinformation.

Are we willing to reimagine our current structures and operations, or are we going to allow disinformation to reign?

Michelle Daniel is the founder and CEO of Connexions Global Matrix, Inc., nonprofit devoted to advancing public education and developing expert, multi-sector, global leadership in the information and technology space. Previously, she served as assistant director at the Center for Russian, East European, & Eurasian Studies at the University of Texas at Austin