

GLOBAL TASK FORCE
**PREDICTIVE
ANALYTICS**
FOR SECURITY
AND DEVELOPMENT



RESPONSIBLE AND SAFE AI:

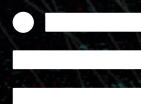
A Primer for Policymakers in the Global South

By Giulia Neaher with Gordon Laforge, Robert Muggah and Gabriella Seiler

This report was funded by the Global Innovation Fund



IGARAPÉ INSTITUTE
a think and do tank



**NEW
AMERICA**

Index

Summary	3
Introduction	4
Part I: Understanding Responsible AI Policy	
Why Responsible AI?	6
Reviewing the AI Lifecycle	7
Assessing Risks of AI	9
Responsible AI Policy Efforts Around the World	12
Part II: Creating Responsible AI Policy	
Task Force Recommendations: Identifying and Mitigating Key AI Risks	15
A Responsible AI Policy Roadmap	17
Closing Reflections	20
Endnotes	21

RESPONSIBLE AND SAFE AI:

A Primer for Policymakers in the Global South

By Giulia Neaher with Gordon Laforge, Robert Muggah and Gabriella Seiler

April 2024

Summary

Artificial intelligence (AI) is rapidly shaping global politics, economies, and societies worldwide. To keep up with the pace of AI innovation and ensure the technology produces benefit without doing harm, policymakers must establish rules and strategies to manage risk and support beneficial development. Many countries are just beginning to define AI policy and regulation. This Primer is an accessible guide to "responsible AI" for policymakers in the Global South. Part one presents background on AI technology, risks, and examples of policies from different jurisdictions around the world. Part two provides recommendations and steps for designing and implementing responsible AI policy. The primer draws on the insights and work of the Global Task Force on Predictive Analytics for Security and Development, an expert group overseen by the Igarapé Institute and New America that came together in 2023 to discuss principles and practices for safe and equitable AI development and deployment.

Introduction

In 2023, artificial intelligence (AI) shot to the top of the global agenda.¹ Technological advances, regulatory efforts, and public hype around generative AI in particular made algorithms and large language models (LLMs) a priority focus for countries and companies around the globe. Highly capable generative AI models such as ChatGPT 4.0, Claude, Gemini, and Bard came to market; and hundreds of billions of dollars in investment poured into established companies and start-ups alike. Looking ahead, this trend is likely to continue - the World Economic Forum's annual summit in Davos named AI a key priority for 2024.²

As the uptake of powerful AI models by governments, firms, and individuals increases, multipurpose, adaptive, and self-learning AI will become a part of daily life.³ Notwithstanding its extraordinary promise, the rise of AI brings with it a variety of technological, security, social, and economic risks. In order to mitigate these risks, policymakers are adopting the paradigm of responsible AI. Broadly, responsible AI refers to AI that is used for the public benefit and has built-in safeguards to prevent societal harms. The Global Partnership on Artificial Intelligence (GPAI) defines responsible AI as “a vision of AI that is human-centered, fair, equitable, inclusive and respectful of human rights and democracy, and that aims at contributing positively to the public good.”⁴

Box 1. The UN's Guiding Principles for Responsible AI

In its Interim Report on Governing AI for Humanity, the United Nations High-Level Advisory Body on AI identifies five principles that should form the basis of responsible AI governance:

- **Inclusivity:** *all citizens, including those in the Global South, should be able to access and meaningfully use AI tools.*
- **Public Interest:** *governance should go beyond the do no harm principle and define a broader accountability framework for companies that build, deploy and control AI, as well as downstream users.*⁵
- **Centrality of data governance:** *AI governance cannot be divorced from the governance of data and the promotion of data commons.*
- **Universal, networked and multistakeholder:** *AI governance should prioritize universal buy-in by countries and stakeholders. It should leverage existing institutions through a networked approach.*
- **International Law:** *AI governance needs to be anchored in the UN Charter, International Human Rights Law, and the Sustainable Development Goals.*

As AI systems become more powerful and pervasive, it will become increasingly important for policymakers to mandate responsible AI principles that ensure individuals and societies are protected from AI risks. Despite some promising signs (see Box 2), such efforts to develop and implement responsible AI principles are lagging behind in the Global South, given the comparative lack of AI uptake, industrial foothold, and local capacity.⁶

Box 2. Notable Responsible AI Efforts in the Global South

In an effort to bridge this Responsible AI gap, policymakers in the Global South are working to encourage domestic AI sector growth and simultaneously take steps to protect against AI risks. As of March 2024, dozens of countries in the Global South have been developing AI strategies and policies. The African Union is expected to release its strategy for AI, which is predicted to suggest policy steps that policymakers in AU countries can undertake domestically and emphasize the importance of protecting Africa from dominance by Western AI companies.⁷ Brazil's legislature will also continue to discuss Bill 2338, which outlines a broad AI regulatory framework and has been the subject of heated debate since it was first proposed in January 2023.⁸

In 2023, the Igarape Institute and New America convened the Global Task Force on Predictive Analytics,⁹ a group of 15 esteemed technologists, ethicists, researchers, and activists from across the Americas, Africa and Asia. Over the course of the year, the Task Force developed principles and recommendations for the responsible design, development and deployment of predictive AI tools in the domains of public safety and sustainable development in the Global South.

Drawing on the Task Force's insights, this Primer presents a primer on ethical AI policy and offers practical recommendations for leaders in lower- and middle-income settings. The Bulletin begins with an overview of responsible AI. It then provides a review of the targets of AI policy -- AI technology itself and AI risks -- before examining global approaches to responsible AI policymaking. The Bulletin concludes with actionable steps and a policy roadmap for decision-makers, practitioners, and advocates.

Part I: Understanding Responsible AI Policy

Why Responsible AI?

Many issues are at play in AI design, development, procurement, and deployment. Among these are data availability and integrity, biases in AI algorithms, labor force impacts of AI, environmental sustainability, economic profitability, public adoption and trust in AI, and others.¹⁰ When evaluating and addressing such policy issues, a responsible AI approach that keeps safety and ethics front of mind is critical.

Box 3. Why Care About Responsible AI?

The Responsible Artificial Intelligence Institute highlights three core reasons why policymakers need to develop responsible AI policy now.¹¹

1. AI is Everywhere: *We should care about responsible AI because AI impacts many of our daily lives. AI is involved when you log onto social media, request a loan from the bank, check into the doctor's office, and travel in airports;*

2. AI is Growing: *This problem is urgent and significant, particularly because AI adoption is widespread. As of 2022, the global AI market is valued at over \$387 billion and is projected to more than triple by 2029; and*

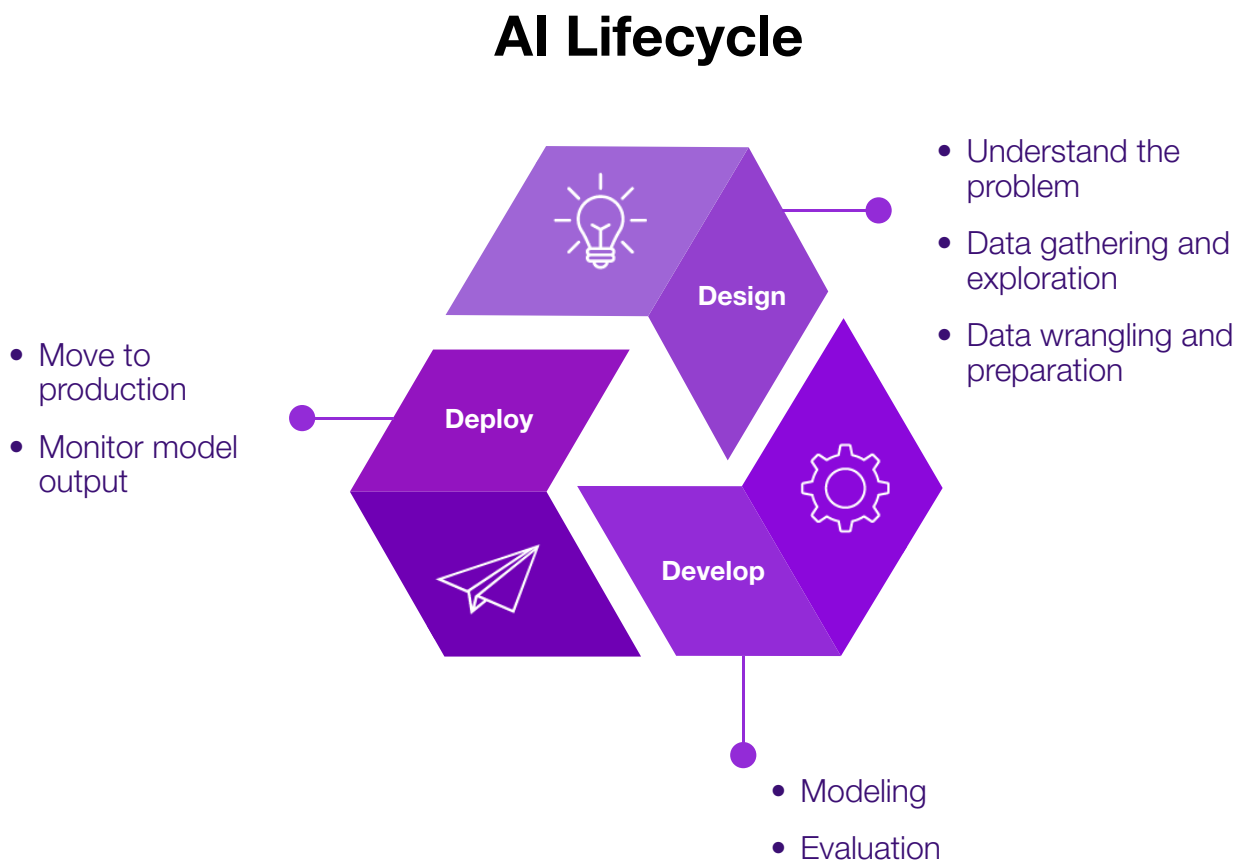
3. AI Can Magnify Harm: *Because AI applications already surround us and AI can be applied easily on a broad scale, it has the potential to amplify the biases of its creator.*

In the private sector, practicing responsible AI from the beginning matters because it will help drive up user trust and adoption, ensure compliance with future government regulation, and protect companies from reputational risk.¹² In the public sector, leaders must pursue responsible AI internally and build regulation that promotes responsible AI that is tailored to the needs and realities of diverse communities including the most vulnerable as part of their duty of care to their constituents. If policymakers do not keep ethics and safety at the forefront of AI policy discussions, they risk inadequately protecting citizens and societies from a technology that is already changing how economies, countries, and people behave around the world. Moreover, failure to protect citizens can also set back innovation – for example, if an AI-driven technology is deployed and injures individuals or causes rapid economic and social dislocations, political and civil backlash will halt AI adoption in its tracks.

Reviewing the AI Lifecycle

A review of responsible AI policies suggests that they typically consider different phases across the ‘AI lifecycle,’ or the process that takes an AI technology from ideation to market.¹³ Understanding this lifecycle is crucial to assess and create AI policy options, and many expert recommendations on AI draw heavily on a simple framework that considers three phases: design, development, and deployment (see Figure 1).

Figure 1. AI lifecycle¹⁴



Source: GSA - IT Modernization Centers of Excellence.

Designing AI: There are at least two key components of the *design* phase: “problem definition” and “data acquisition”. In the former, AI developers come up with an idea for a new model -- they identify a problem they would like to solve with AI and decide how to do it. Once developers have an AI model in mind, they set the stage to build the model by collecting necessary training data (see Box 4).

Box 4. What is Training Data?

AI models are built by “training” algorithms on massive quantities of data. AIs can be categorized by the ways they interact with this data -- most commonly, researchers split models into “supervised” or “unsupervised” AI.¹⁵

In “supervised” AI, individual data points are categorized and labeled for algorithms to read; the algorithms then sort through the data and attempt to correctly classify them according to the predetermined labels. Developers can give models direct feedback by rating the quality of their classifications. By iterating through this process over and over again, AI models can “learn” to correctly identify unlabeled data points outside of the original dataset.¹⁶

In “unsupervised” AI, algorithms sort through data on their own, using advanced statistical analysis to find similarities or links between data points. This is the method utilized by generative AI models like ChatGPT, which creates responses to inquiries by looking for patterns amongst vast quantities of training data.¹⁷ Developers give feedback on the results produced by AIs but have little insight into or impact on how the AI obtains those outputs.

Development: During the *development* phase, AI models are built and evaluated.¹⁸ Algorithms are written and trained iteratively on the datasets mentioned above, and researchers evaluate the quality of outputs to determine whether and how the algorithms need to be adjusted. Once the researchers are happy with the performance of the model, the AI moves into the deployment stage.

Deployment: Throughout the *deployment* stage, the AI is put into production and brought to users and to market. It can be sold or licensed to institutions and consumers, who input their own data and use the AI for the function identified in the *design* stage. In the *deployment* stage, the developers of the AI monitor its functionality to ensure it is working as intended.¹⁹

Assessing Risks of AI

AI risks are the potential harmful effects generated by AI at any point during the lifecycle. Though AI risk assessments vary, there are a few that are widely recognized in most AI policy frameworks (see Figure 2).

Figure 2. A sample of AI-related risks



Problems with transparency, explainability, and accountability:

A key challenge for AI policymakers is the difficulty of understanding how they actually work. Many companies want to keep their algorithms, datasets, and training methods secret in order to maintain a competitive edge in the market; in 2023, for instance, OpenAI faced huge controversies over its secretive business practices.²⁰ This lack of transparency is further complicated by the so-called “black box problem,” which refers to the fact that AI developers themselves often do not understand the inner workings of unsupervised AI models. These models discover patterns in data on their own, and developers can only evaluate the AIs’ results, not how they get there.²¹ Without insight into how AI models make decisions, policymakers face enormous difficulty in holding developers accountable for incorrect or harmful outputs that models may produce.



Cybersecurity:

There are a wide variety of cybersecurity risks in today’s AI models, and the UK National Cyber Security Centre recently predicted that AI will dramatically increase the volume and speed of cyber attacks.²² In particular, many experts fear “adversarial” AI attacks, in which hackers manipulate AI systems by muddling their algorithms and data. Such attacks could potentially take critical AI-driven systems out of commission.²³ The speed of AI deployment around the world may further exacerbate cybersecurity risks; in a recent statement, the U.S. National Institute of Standards and Technology (NIST) noted that many organizations are implementing AI without adequately putting AI-specific cybersecurity mechanisms into place.²⁴



Bias and discrimination:

Embedded bias in datasets and algorithms used by unsupervised AI models like ChatGPT is a significant civil rights risk. For example, if an AI detects a pattern in which women are associated with certain professions and men with others, then the responses or decisions of that AI are likely to replicate that pattern.²⁵ As a result, the AI outputs may perpetuate societal stereotypes about the jobs available to men and women. For this reason, the selection of datasets used to train AIs can be instrumental in embedding bias into that AI’s results; if a dataset contains too much gendered language, the AI will likely replicate that language. Bias can also come from poor training or algorithmic design on the part of developers.



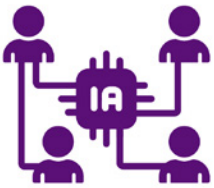
Intellectual property violations:

Intellectual property problems have rapidly gained attention in the last year, particularly as generative AI models capable of making high-quality images have improved. One problem arises from the fact that the training data used for generative AI models may include copyrighted content or images, often used without consent of copyright owners.²⁶ This use of copyrighted images can then create more problems by leading AIs to produce outputs that are similar or nearly identical to copyrighted intellectual property.²⁷



Privacy violations:

Privacy concerns regarding AI are broad. Training datasets may also contain private personal information, which may threaten individual privacy rights, especially if AI outputs reveal confidential information. Specific applications of AI may also harm privacy rights; AI-powered facial recognition technology, in particular, has become a source of controversy in recent years.²⁸



Adverse societal effects:

AI experts also highlight many ways that AI could directly harm parts of society. For example, AI could make certain professions obsolete and displace many workers. Floods of AI-generated disinformation could weaken trust in societies. And environmental sustainability is also a problem; AI requires a huge amount of computational power to train and operate, which in turn relies on significant energy and water consumption.²⁹



Misalignment:

Some experts also highlight the potential for AI to have far-reaching, unintended impacts in the long term, if AI systems' emergent capabilities lead them to “go rogue” and act in harmful ways outside of the scope intended by their creators.³⁰ However, the importance of these risks is hotly debated, with many researchers stating that focusing on long-term risks distracts from the problems that are already affecting people today.³¹ At present, these risks are largely hypothetical.

The aforementioned list of risks is not comprehensive. Different organizations, researchers, and developers all have their own risk frameworks. There are competing schools of thought around which AI risks are most dangerous or urgent, and new risks arise as the technology itself evolves.³²

Though often starting with risks, many responsible AI policies and frameworks orient around principles and solutions. For example, rather than name threats to personal privacy, an AI policy might espouse the right to privacy or a commitment to privacy protection.³³ This broader principles-based approach is helpful given the constantly changing nature of AI technology and risk. By avoiding too much specificity and focusing on rights and outcomes rather than technical specifications, these principles have a greater chance of holding up amid future technological developments.

Box 5. UNESCO'S APPROACH TO HUMAN RIGHTS-CENTERED AI³⁴

UNESCO'S Guideline on Ethical AI proposes not only a set of principles for AI development, but also high-level values to guide the AI sector as a whole. This broader, human rights driven approach aligns with the organizational character of UNESCO, illustrating the way that AI risk assessment is also a product of organizational (or national) context.

Values:

- *Respect, protection and promotion of human rights and fundamental freedoms and human dignity*
- *Environment and ecosystem flourishing*
- *Ensuring diversity and inclusiveness*
- *Living in peaceful, just and interconnected societies*

Principles:

- *Proportionality and Do No Harm*
- *Safety and Security*
- *Fairness and non-discrimination*
- *Sustainability*
- *Right to Privacy, and Data Protection*
- *Human oversight and determination*
- *Transparency and explainability*
- *Responsibility and accountability*
- *Awareness and literacy*
- *Multi-stakeholder and adaptive governance and collaboration*

Responsible AI Policy Efforts Around the World

Countries and governments tend to align national priorities with their approaches to AI regulation. As a result, the flavor of ethical AI policies issued by current leaders in AI regulation - namely, the U.S., EU, and China - are very different.

The approach taken by **the U.S.** (and by allies in the UK and Japan) has been to issue guidance regarding AI risks but to mostly hold off from concrete regulation; by shying away from restrictive regulations on the private sector, the U.S. hopes to encourage domestic innovation and leadership in AI on the world stage.³⁵ In practice, this translates into the construction of a patchwork of AI regulations, stronger in some areas than in others.³⁶ For example, the 2023 U.S. Executive Order on AI places many restrictions on *government* use and development of AI, but largely forgoes binding rules for the private sector.³⁷

The EU has taken a much stronger stance on regulation of the private sector, by not only issuing high-level principles but also approving the world's first comprehensive AI policy framework. This framework, the EU AI Act passed into law in March 2024 and will establish an industry-wide set of regulations. These regulations place an emphasis on the *Deployment* stage but may affect all stages of the lifecycle; they categorize AI applications by level of risk to society and place additional upstream regulations on higher-risk applications.³⁸

Meanwhile, in China, policymakers have issued binding interim AI regulations that aim to ensure development of the industry but also guarantee state supervision of AI and compliance with national core values. The Chinese regulations mostly target the *Development* phase of AI by setting requirements for algorithms used to build AI models.³⁹ Rules and regulations continue to roll out, and experts do not expect to see a comprehensive draft AI law until late 2024.⁴⁰

Table 1. Comparing two approaches to comprehensive responsible AI policy: U.S. vs. EU

Policy	<i>The White House Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (U.S.)</i>	<i>The EU AI Act (EU)</i>
Description	<i>In October 2023, the U.S. issued its most significant piece of AI policy to date: the White House Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.⁴¹ This regulation makes a strong commitment to safe and ethical AI, anchored throughout the AI lifecycle, and it contains over 100 pages of AI principles and regulations. A few key policy initiatives and instruments are outlined below.</i>	<i>The EU AI Act's chief innovation is that of its risk framework, in which AI applications are categorized according to the level of risk they pose to the rest of society. AI applications deemed to be too risky are entirely prohibited, while high-risk applications are subject to a myriad of regulations. Lower-risk applications correspondingly face less regulation.</i>
Key Initiatives and Instruments	<ul style="list-style-type: none"> - Development of standards and principles for safety, security, and trustworthiness. <i>The U.S. National Institute of Standards and Technology (NIST) was tasked with creating a system of standards, tools, and tests to be used by the U.S. government when adopting and deploying AI technologies, intended to ensure that such uses of AI are safe. NIST was also tasked with creating guidelines and principles (less binding than mandated standards) designed to inform industry practices with regard to trustworthy AI.⁴²</i> - Division of labor between government agencies. <i>The Executive Order splits up its many mandates among various government agencies, which avoids putting too much power or responsibility with one branch of government and encourages a holistic, all-of-government approach. It also ensures that issues will be assigned to the agencies best equipped (whether in terms of funding, staffing, or expertise) to address them.⁴³</i> 	<ul style="list-style-type: none"> - Tiered categorization and regulation of AI risks. <i>The Act splits AI systems into four risk levels: unacceptable, high, limited, and minimal risk.⁴⁴ Each level faces a different regulatory regime. "Unacceptable" AI applications include biometric identification, social scoring, and other uses deemed to directly infringe upon civil rights. "High-risk" applications include those that relate to energy infrastructure, medicine, and criminal justice.</i> - Regulatory rules for "foundation models." <i>Foundation models are general-purpose AIs (including chatbots like ChatGPT) that can be licensed or used across a wide variety of applications. Under the AI Act, providers of such AIs will have to adhere to some baseline safety standards and may face additional regulation if their models are deemed to bear "systemic risk."⁴⁵</i>

Countries and companies across the Global South are also working to develop AI policy, sometimes with differing priorities from the Global North. One such concern is economic growth; current projections say that only a small portion of global financial benefits from AI will go to the developing world.⁴⁶ Other key concerns for the Global South are shrinking the digital divide, capacity building, and supporting domestic AI markets when discussing AI policy.

The African Union is adopting this approach, as shown in a press release from its November meeting that states:

AI has the potential to significantly impact the attainment of Agenda 2063 and the Sustainable Development Goals (SDGs). AI is important to Africa because of its economic, social, political and geopolitical impact. AI technologies can stimulate economic growth by creating new industries, driving innovation, and generating employment opportunities.⁴⁷

Box 6. A Sample of AI Priorities From the Global South.

ECONOMIC GROWTH. Policies supporting economic growth may focus on:

- Capacity-building. This could look like digital skill-building and education, creation of incentives to retain domestic AI talent (and prevent “brain drain”), and furthering efforts to increase digital penetration.
- Supporting local AI ecosystems. Though there is a growing community of AI startups based in the developing world, there is also a much smaller market for AI products. Recent data from IBM shows that while 60% of Chinese companies currently use some form of AI, only 7% of companies from the Global South do so.⁴⁸ Governments may choose to enact policy that financially or otherwise supports local AI markets.

EQUITY. Leaders in the Global South are also concerned about equitable distribution of AI gains around the globe. Policies furthering equity may focus on:

- Building datasets that are representative of the Global South. Low digital penetration in the Global South results in a dearth of training data from the developing world, which in turn means that the datasets used to build major AI models are trained on data that is biased towards advanced economies, where technology use is more prevalent.⁴⁹ To mitigate this, policymakers may choose to support the collection of domestic or regional datasets.
- Setting international standards for responsible AI that are representative of the Global South. Much international policymaking in responsible AI has stemmed from the Global North. Global South policymakers may choose to issue their own regional responsible AI commitments or to seek increased leadership in AI policy via international organizations.

Part II: Creating Responsible AI Policy

Task Force Recommendations: Identifying and Mitigating Key AI Risks

The Global Task Force on Predictive Analytics developed a number of recommendations for the design and deployment of responsible AI, specifically in the Global South. A selection of key recommendations is outlined below, and the full set of recommendations can be found in the Task Force's Technical Notes.⁵⁰

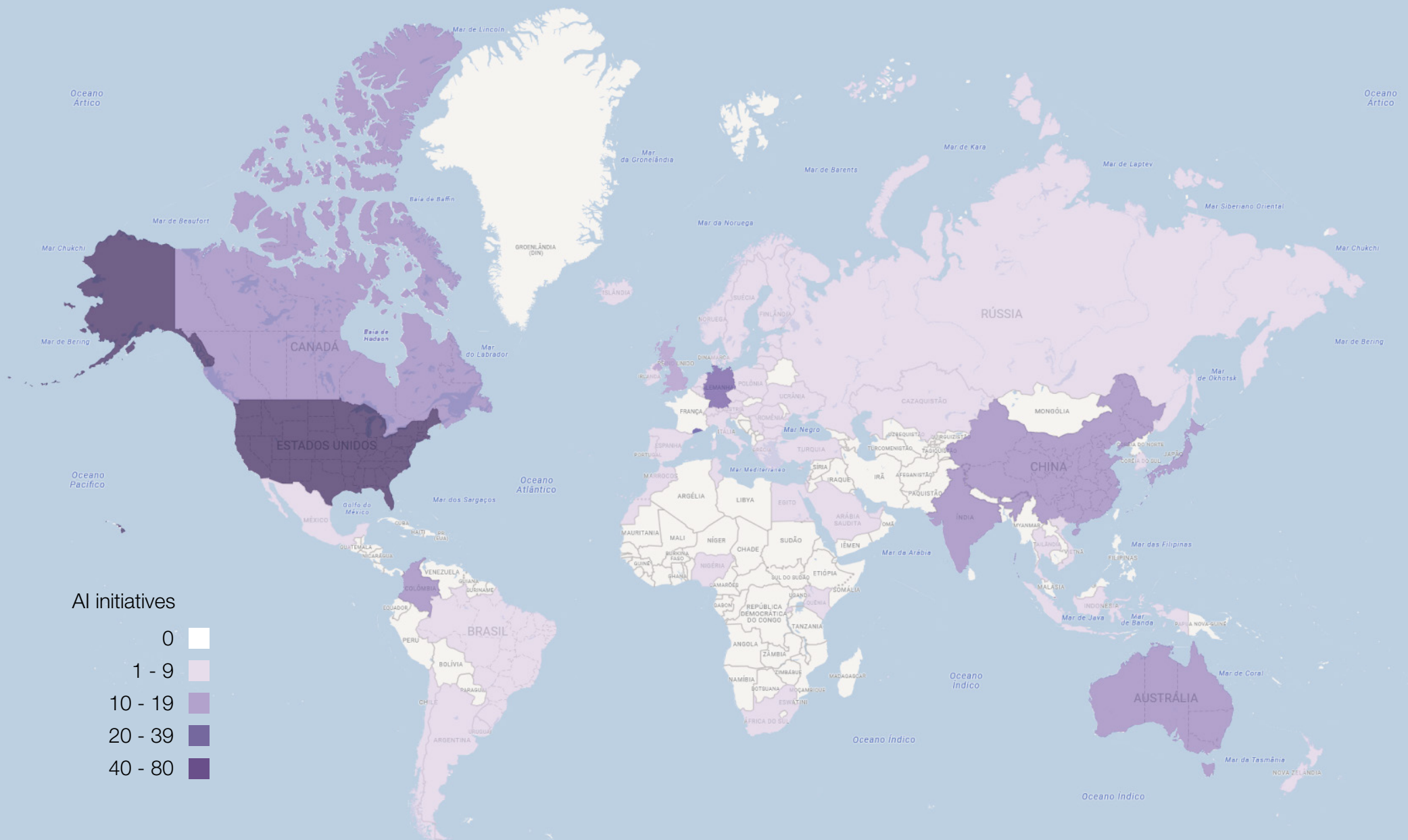
The Task Force suggests that responsible **design and development of AI** should be:

- **Fair.** Algorithms should be designed to treat all users equally.
- **Participatory.** AI design should have input from diverse groups that include civil society members, researchers, and policymakers.
- **Representative.** Datasets and algorithms should be unbiased and inclusive.
- **Interpretable.** AI systems should be understandable to the average user, and AI creators should be clear and open about their design and development processes.
- **Accountable.** Mechanisms should be set to hold AI creators, users, and other industry actors responsible for AI harms as needed.

Deployment. The Task Force found that **responsible AI deployment** will require:

- *Informational symmetry.* Purchasers or users of AI may not fully understand the costs and benefits of AI technology, particularly in the Global South, where access to information about AI technology suppliers may be more limited. Data sharing and due diligence mechanisms to increase access to information may increase informational symmetry.
- *Data practices that are tailored to lower-income settings.* Many AIs are trained on data coming from the Global North; in order for the technology to work accurately, it will be necessary to collect representative datasets for the Global South.
- *Safeguards against intentional and unintentional misuse of AI.* For instance, there are risks associated with using AI for law enforcement, as basing arrests or legal decisions on AI outputs, which can be inaccurate, may imperil innocent individuals. Safeguards are necessary to protect from these and other harms.
- *Continuous updating of AI tools to ensure compliance with laws.* AI technology is rapidly evolving, as are the laws surrounding it. It will be necessary to build flexibility into AI regulatory regimes to prepare for future adjustments.
- *Strong monitoring and evaluation practices.* After deployment, AI should be subjected to transparent, inclusive, iterative, and interpretable monitoring and evaluation.

Figure 3: Density of safe and ethical AI policies and guidance



A Responsible AI Policy Roadmap

This section weaves the foundations outlined above into a brief how-to guide for leaders interested in creating AI policy in the Global South. This guide is not intended to be comprehensive; it aims to give policymakers, organizations, and advocates in the Global South a good starting point from which to build their own AI policies or strategies.

Designing AI Policy

If you are interested in creating an AI policy, the first step is to study the problem you are trying to solve. What are you trying to do, and how do you plan to do it?

Consider first the *contents* of your policy:

- Do I want to prioritize capacity-building? How will a potential policy affect my country's AI sector?
- Is there a specific AI risk I want to address? Which risks are important to me or my country?
- What phase of the AI lifecycle would my policy target? Do I want to address datasets, algorithms, or applications in a specific policy?

And also consider the *structure* of your AI policy:

- What policy avenues are available to me? Do I want to create binding legislation or a recommendation? What challenges face me if I pursue either or both of those routes?
- If I want to build legislation, do I want it to take a vertical or horizontal approach? How broad should the scope of my policy be?

As with any policy, the design phase will also be crucial for identifying legislative pathways you want to pursue, stakeholders you want to recruit or avoid, budgetary options, and other relevant factors (see Box 7).

Table 2. A sample of ethical AI policies

This graphic draws on safe and ethical AI principles from around the world and presents some specific policy interventions that a policymaker might choose to pursue in relation to their goals.

Principle or Goal	Policy Intervention
<i>Transparency, explainability, and accountability</i>	<ul style="list-style-type: none"> - <i>Require the sharing of datasets or algorithms with government auditors or with the public</i> - <i>Establish who is accountable for specific AI harms</i>
<i>Bias and discrimination</i>	<ul style="list-style-type: none"> - <i>Set standards to prevent algorithmic and data bias and discrimination</i> - <i>Monitor or restrict the use of AI in high-risk applications, such as law enforcement</i>
<i>Intellectual property violations</i>	<ul style="list-style-type: none"> - <i>Introduce content authentication or watermarking to label AI-generated content</i> - <i>Set rules for the use of copyrighted content in training AIs</i>
<i>Privacy</i>	<ul style="list-style-type: none"> - <i>Review data collection practices (both by AIs themselves and for AI training datasets)</i> - <i>Reinforce or develop new cryptographic tools to protect individual data</i> - <i>Set standards for the collection and use of personal data related to AI</i> - <i>Limit use of AI for facial recognition or other biometric recognition</i>
<i>Human rights</i>	<ul style="list-style-type: none"> - <i>Enforce standards preventing data or algorithmic bias</i> - <i>Establish best practices for AI use in the criminal justice system, healthcare, and education</i>
<i>Labor rights</i>	<ul style="list-style-type: none"> - <i>Create labor laws protecting both AI workers and people who may lose their jobs due to AI</i> - <i>Train workers to enter the AI sector</i>
<i>Environmental impacts</i>	<ul style="list-style-type: none"> - <i>Monitor or limit energy consumption by AI data centers</i> - <i>Secure supply chains for critical minerals used in AI</i>
<i>Cybersecurity</i>	<ul style="list-style-type: none"> - <i>Strengthen cryptographic tools for AI data and models</i> - <i>Improve cybersecurity practices for AI development and deployment</i> - <i>Conduct red-team testing to look for AI vulnerabilities</i>
<i>Government leadership in AI</i>	<ul style="list-style-type: none"> - <i>Issue national guidance for safe and ethical AI</i> - <i>Improve AI practices within government (preventing bias, developing safety standards, etc.)</i> - <i>Accelerate recruitment of AI professionals in government</i> - <i>Develop regulatory incentives including sandboxes to foster AI experimentation and testing prior to deployment</i> - <i>Conduct government-sponsored research on AI technology</i>
<i>National security</i>	<ul style="list-style-type: none"> - <i>Require AI developers to share safety and test results with government</i> - <i>Develop safety standards before deployment</i> - <i>Prohibit the use of AI in developing dangerous biological materials</i>
<i>Economic growth</i>	<ul style="list-style-type: none"> - <i>Invest in innovation and AI sector growth</i> - <i>Disincentivize “brain drain” of national AI talent</i> - <i>Support the adoption of AI across industries like healthcare and agriculture</i>

Developing AI Policy

After designing the idea and goals for an AI policy, you will enter the development phase. This will be largely context-driven and will depend on the organizational or national structure within which you operate. Policy development may involve stakeholder consultations, legislative voting and negotiation, or any number of other policy techniques. A step-by-step approach to AI policy development could include:

1. Conduct internal AI policy goal setting and design
 - a. Hold internal meetings to identify key political or policy priorities, stakeholders, and leaders
2. Assign responsibility for management and leadership of AI policy design project
3. Establish a formal AI taskforce of selected experts, tasked with studying potential AI policies
 - a. Assign this selected group to produce a key deliverable; this could be a draft piece of legislation, a recommendation for the president or executive branch, or a roadmap for future pursuit of either of those avenues
4. Hold expert consultations to evaluate the findings and recommendations of the AI taskforce
 - a. These consultations should include the private sector, to ensure that any policy decisions are feasible and will be reliably implemented by AI companies
5. (Optional) Return AI policy recommendation to taskforce to review
6. Act on AI policy recommendation

Box 8. Brazil's AI Policy: Design and Development

Brazil has taken significant steps towards regulating AI domestically in recent years, but the country's path through the Development stage has been difficult. Brazil first entered the Design stage in 2022 with the establishment of a legal commission mandated to study potential methods for AI regulation.⁵¹ The commission's findings were used to draft Bill No.2338/2023 (the Draft AI Law), and that bill was introduced to the Senate in January 2023.⁵² The Draft AI Law emulates the risk-based approach taken by the EU AI Act but also includes language on individual civil rights.⁵³ After nearly a year of debate, Senator Astronaut Marcos Pontes introduced a new version of the bill in December 2023, which reduces the emphasis on individual rights and loosens regulation with the goal of encouraging innovation.⁵⁴ Further iterations of the bill are expected in 2024.

Deploying AI Policy

In this stage, the AI policy is disseminated publicly. Policymakers or issuing agencies should conduct regular monitoring of the policy and of the state of AI writ large; AI is a rapidly developing technology, and technological advances will necessitate adjustments to existing AI guidelines. Monitoring may be led by implementing agencies, advisory bodies, or by advocates who work to hold companies and governments accountable. AI policy should be iterative due to the speed of technological growth and global adoption - what works today may not work as well tomorrow.

Closing Reflections

AI is increasingly a central feature of daily life for hundreds of millions of people around the world. It will soon be used by billions in trillions of daily actions. It is clear that both its risks and benefits will have widespread impact on societies around the world. It is vital that the Global South engage with AI, in order to ensure that the economic gains of AI contribute to development goals and that communities are protected from AI harms. In the coming months and years, leaders in the Global South will need to quickly but thoughtfully evaluate their approaches to AI governance. This note may serve as a useful guide of practical recommendations for AI policymakers. Ultimately, AI policy should be informed by the needs and benefits of users, the risks associated with the technology, and wider social and economic aspirations. By approaching AI policy systematically and holistically, leaders can effectively contribute to AI governance that serves their regions, countries, and communities.

Endnotes

- 1_ McKinsey & Company. "The State of AI in 2023: Generative AI's Breakout Year." QuantumBlack AI by McKinsey, 1 Aug. 2023, <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year>.
- 2_ Whiting, Kate. "From Sam Altman to António Guterres: Here's What 10 Leaders Said about AI at Davos 2024." World Economic Forum, 23 Jan. 2024, <https://www.weforum.org/agenda/2024/01/what-leaders-said-about-ai-at-davos-2024/>.
- 3_ PricewaterhouseCoopers. "2024 AI Business Predictions." PwC, <https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-predictions.html>.
- 4_ "Responsible AI - GPAI." The Global Partnership on Artificial Intelligence, <https://gpai.ai/projects/responsible-ai/>.
- 5_ United Nations. "AI Advisory Body." United Nations, United Nations, <https://www.un.org/en/ai-advisory-body>.
- 6_ Neaher, Giulia, et al. Global Futures Bulletin: Responsible Artificial Intelligence Efforts In the Global South. Instituto Igarapé, 29 Jan. 2024, <https://igarape.org.br/en/global-futures-bulletin-responsible-artificial-intelligence-efforts-in-the-global-south/>.
- 7_ Ryan-Mosley, Tate, et al. "What's next for AI Regulation in 2024?" MIT Technology Review, 5 Jan. 2024, <https://www.technologyreview.com/2024/01/05/1086203/whats-next-ai-regulation-2024/>.
- 8_ Mendes, Laura Schertel, and Beatriz Kira. "The Road to Regulation of Artificial Intelligence: The Brazilian Experience." Internet Policy Review, 21 Dec. 2023, <https://policyreview.info/articles/news/road-regulation-artificial-intelligence-brazilian-experience/1737>.
- 9_ <https://igarape.org.br/en/technical-notes-global-task-force-on-predictive-analytics-for-security-and-development/>
- 10_ Klaus, Ian, and Ben Polsky. "Subnational Practices in AI Policy: A Working Guide." Carnegie Endowment for International Peace, 12 Dec. 2023, <https://carnegieendowment.org/2023/12/12/subnational-practices-in-ai-policy-working-guide-pub-91186>.
- 11_ Lawson, Amanda. "AI vs. Responsible AI: Why Is It Important?" Responsible AI, 24 Jan. 2023, <https://www.responsible.ai/ai-vs-responsible-ai-why-is-it-important/>
- 12_ Mills, Steven, and Daniel Lim. "This Is Why We Need to Talk about Responsible AI." World Economic Forum, 12 Nov. 2020, <https://www.weforum.org/agenda/2020/11/this-is-why-we-need-to-talk-about-responsible-ai/>.
- 13_ "Understanding and Managing the AI Lifecycle." GSA - IT Modernization Centers of Excellence, <https://coe.gsa.gov/coe/ai-guide-for-government/understanding-managing-ai-lifecycle/>.
- 14_ Ibid.
- 15_ Desouza, Kevin C., et al. "Designing, Developing, and Deploying Artificial Intelligence Systems: Lessons from and for the Public Sector." Business Horizons, vol. 63, no. 2, Mar. 2020, pp. 205–13. www.sciencedirect.com.ezp-prod1.hul.harvard.edu, <https://doi.org/10.1016/j.bushor.2019.11.004>.
- 16_ Jacoby, KJ. "Tech Explainer: What Is AI Training?" Performance Intensive Computing, 10 Mar. 2023, <https://www.performance-intensive-computing.com/objectives/tech-explainer-what-is-ai-training>.
- 17_ Gewirtz, David. "How Does ChatGPT Actually Work?" ZDNET, 20 Sept. 2023, <https://www.zdnet.com/article/how-does-chatgpt-work/>.
- 18_ "Understanding and Managing the AI Lifecycle." GSA - IT Modernization Centers of Excellence.
- 19_ Ibid.
- 20_ Dave, Paresh. "OpenAI Quietly Scrapped a Promise to Disclose Key Documents to the Public." WIRED, 24 Jan. 2024, <https://www.wired.com/story/openai-scrapped-promise-disclose-key-documents/>.
- 21_ "Cracking the Code: The Black Box Problem of AI." ScaDS.AI, 19 July 2023, <https://scads.ai/cracking-the-code-the-black-box-problem-of-ai/>.
- 22_ "The Near-Term Impact of AI on the Cyber Threat." UK National Cyber Security Centre, 24 Jan. 2024, <https://www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat>.
- 23_ Dempsey, Jim. "Addressing the Security Risks of AI." Lawfare, 11 Apr. 2023, <https://www.lawfaremedia.org/article/addressing-security-risks-ai>.
- 24_ "NIST Warns of Security and Privacy Risks from Rapid AI System Deployment." The Hacker News, 8 Jan. 2024, <https://thehackernews.com/2024/01/nist-warns-of-security-and-privacy.html>.
- 25_ Packer, Ben, et al. "Text Embedding Models Contain Bias. Here's Why That Matters." Google for Developers, 13 Apr. 2018, <https://developers.googleblog.com/2018/04/text-embedding-models-contain-bias.html>.
- 26_ Appel, Gil, et al. "Generative AI Has an Intellectual Property Problem." Harvard Business Review, 7 Apr. 2023. hbr.org, <https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem>.
- 27_ Zirpoli, Christopher T. Generative Artificial Intelligence and Copyright Law. Legal Sidebar, LSB10922, Congressional Research Service, 29 Sept. 2023, <https://crsreports.congress.gov/product/pdf/LSB/LSB10922>.
- 28_ Sher, Gai, and Ariela Benclouch. "The Privacy Paradox with AI | Reuters." Reuters, 31 Oct. 2023, <https://www.reuters.com/legal/legalindustry/privacy-paradox-with-ai-2023-10-31/>.
- 29_ Hanna, Alex, and Emily M. Bender. "AI Causes Real Harm. Let's Focus on That over the End-of-Humanity Hype." Scientific American, 1 Feb. 2024, <https://www.scientificamerican.com/article/we-need-to-focus-on-ais-real-harms-not-imaginary-existential-risks/>.
- 30_ "AI Risks That Could Lead to Catastrophe." Center for AI Safety, 9 Oct. 2023, <https://www.safe.ai/ai-risk>.
- 31_ Hanna, Alex, and Emily M. Bender. "AI Causes Real Harm. Let's Focus on That over the End-of-Humanity Hype."
- 32_ Schneider, Bruce, and Nathan Sanders. "Opinion | The A.I. Wars Have Three Factions, and They All Crave Power." The New York Times, 28 Sept. 2023. NYTimes.com, <https://www.nytimes.com/2023/09/28/opinion/ai-safety-ethics-effective.html>.
- 33_ "NIST AIRC - AI Risks and Trustworthiness." National Institute of Standards and Technology, US Department of Commerce, https://airc.nist.gov/AI_RM_F_Knowledge_Base/AI_RM_F_Foundational_Information/3-sec-characteristics.

- 34_ "Ethics of Artificial Intelligence." UNESCO, <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>.
- 35_ Whyman, Bill. "AI Regulation Is Coming- What Is the Likely Outcome?" Center for Strategic and International Studies, 10 Oct. 2023, <https://www.csis.org/blogs/strategic-technologies-blog/ai-regulation-coming-what-likely-outcome>.
- 36_ Ibid.
- 37_ The Executive Order requires the Department of Commerce create a mechanism whereby developers of certain qualifying AI models must share information with the U.S. government. It also requires that the Department of Commerce create a regulatory structure for the resale of American AI internationally.
- See "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence." *The White House*, 30 Oct. 2023, <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>.
- 38_ Engler, Alex. "The EU and U.S. Diverge on AI Regulation: A Transatlantic Comparison and Steps to Alignment." Brookings Institution, <https://www.brookings.edu/articles/the-eu-and-us-diverge-on-ai-regulation-a-transatlantic-comparison-and-steps-to-alignment/>.
- 39_ Sheehan, Matt. "China's AI Regulations and How They Get Made." Carnegie Endowment for International Peace, 10 July 2023, <https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117>.
- 40_ Yang, Zeyi. "Four Things to Know about China's New AI Rules in 2024." MIT Technology Review, 17 Jan. 2024, <https://www.technologyreview.com/2024/01/17/1086704/china-ai-regulation-changes-2024/>.
- 41_ "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence." The White House.
- 42_ Trivedi, Prem M. "Unpacking the White House's Executive Order on AI." New America, 10 Nov. 2023, <http://newamerica.org/oti/blog/unpacking-the-white-houses-executive-order-on-ai/>.
- 43_ Ryan-Mosley, Tate, and Melissa Heikkilä. "Three Things to Know about the White House's Executive Order on AI." MIT Technology Review, 30 Oct. 2023, <https://www.technologyreview.com/2023/10/30/1082678/three-things-to-know-about-the-white-houses-executive-order-on-ai/>.
- 44_ Hoffmann, Tessa. "The EU AI Act: A Primer." Center for Security and Emerging Technology, 26 Sept. 2023, <https://cset.georgetown.edu/article/the-eu-ai-act-a-primer/>.
- 45_ Hainsdorf, Clara, et al. "Dawn of the EU's AI Act: Political Agreement Reached on World's First Comprehensive Horizontal AI Regulation." White & Case LLP, 14 Dec. 2023, <https://www.whitecase.com/insight-alert/dawn-eus-ai-act-political-agreement-reached-worlds-first-comprehensive-horizontal-ai>.
- 46_ Okolo, Chinasa T. "AI in the Global South: Opportunities and Challenges towards More Inclusive Governance." Brookings, 1 Nov. 2023, <https://www.brookings.edu/articles/ai-in-the-global-south-opportunities-and-challenges-towards-more-inclusive-governance/>.
- 47_ "The 5th Ordinary Session of the Specialized Technical Committee on Communication and Information Communications Technology." African Union, 23 Nov. 2023, <https://au.int/en/5thstccict>.
- 48_ Candelon, François, et al. "How A.I. Can Make the Global South More Competitive." Fortune, 6 Jan. 2023, <https://fortune.com/2023/01/06/artificial-intelligence-ai-global-south-national-champions/>.
- 49_ Munga, Jane. "Bridging or Widening the Digital Divide: The Challenge of AI in Africa." Hello Future, 10 Mar. 2023, <https://hellofuture.orange.com/en/bridging-or-widening-the-digital-divide-the-challenge-of-ai-in-africa/>.
- 50_ Muggah, Robert, et al. "Technical Notes: Global Task Force on Predictive Analytics for Security and Development." Instituto Igarapé, 6 Feb. 2024, <https://igarape.org.br/en/technical-notes-global-task-force-on-predictive-analytics-for-security-and-development/>.
- 51_ Oberschelp de Meneses, Anna, et al. "Brazil's Senate Committee Publishes AI Report and Draft AI Law." Inside Privacy, 27 Jan. 2023, <https://www.insideprivacy.com/emerging-technologies/brazils-senate-committee-publishes-ai-report-and-draft-ai-law/>.
- 52_ Ibid.
- 53_ Mendes, Laura Schertel, and Beatriz Kira. "The Road to Regulation of Artificial Intelligence: The Brazilian Experience."
- 54_ Ibid.

Task Force Members



Reuben Abraham

CEO, Artha Global



Yolanda Martinez

GovStack Initiative Lead, International Telecommunications Union



Chinmayi Arun

Research Scholar in Law and Executive Director of the Information Society Project, Yale Law School



Nanjala Nyabola

Writer, researcher, political analyst, and activist



Wafa Ben-Hassine

Human rights lawyer, and Principal, Responsible Technology, Omidyar Network



Taylor Owen

Associate Professor and Founding Director of the Center of Media, Technology and Democracy, Max Bell School of Public Policy, at McGill University



Tiffiniy Cheng

Co-founder, Fight for Future and Open Congress



Maria A. Ressa

CEO, President, and Co-Founder, Rappler and Nobel Laureate



Rumman Chowdhury

Former Director for ML Ethics, Twitter and AI Fellow, Berkman Klein



Fabro Steibel

Executive Director, Institute for Technology and Society Rio



Haksoo Ko

Professor of Law, Seoul National University School of Law (observer)



Stefaan Verhulst

Co-Founder and Chief Research and Development Officer, Governance Lab (GovLab), New York University



Ronaldo Lemos

Professor, Rio de Janeiro State University Law School



Andrew Wyckoff

Director of the Directorate for Science, Technology and Innovation, Organization for Economic Cooperation and Development (OECD)



Vukosi Marivate

ABSA Chair of Data Science, University of Pretoria



IGARAPÉ INSTITUTE a think and do tank

The Igarapé Institute is an independent think and do tank committed to citizen, digital, and climate security. Founded in 2010 and based in Brazil, the Institute designs and delivers data-driven and evidence-based solutions to address global challenges. The Institute oversees activities in over 20 countries across the Americas, Africa and Europe leveraging a combination of primary research, technology innovation, strategic partnerships, and global advocacy and communications. Ranked among the top social policy and environment think tanks in the world and a leading non-profit in Brazil, the Institute works with governments, the private sector, and civil society to deliver lasting social impact.

See www.igarape.org.br



New America is a think tank and civic enterprise dedicated to renewing, reimagining and realizing the Promise of America in an era of rapid technological and social change. Since 1999, New America has nurtured a new generation of policy experts and intellectuals and pioneered bold and successful policy initiatives in areas ranging from education to open technology to political reform. As the nation's premier idea incubator and accelerator, New America has the unique capacity to quickly innovate new programs and scale existing programs that address the most relevant problems in our society — today, and in decades to come.

See www.newamerica.org

www.igarape.org.br



IGARAPÉ INSTITUTE
a think and do tank