Global Trends in AI Governance

Evolving Country Approaches

WORLD BANK GROUP

ABNASIA.ORG

WORLD BANK GROUP

© 2024 International Bank for Reconstruction and Development

The World Bank Group
1818 H Street NW Washington DC 20433
Telephone: 202-473-1000
Internet: www.worldbank.org

DISCLAIMER

This work is a product of the staff of The World Bank with external contributions. The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent.

The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

RIGHTS AND PERMISSIONS

The material in this work is subject to copyright. Because the World Bank encourages dissemination of its knowledge, this work may be reproduced, in whole or in part, for noncommercial purposes as long as full attribution to this work is given.

Any queries on rights and licenses, including subsidiary rights, should be addressed to the Office of the Publisher, The World Bank, 1818 H Street NW, Washington, DC 20433, USA; fax: 202-522-2422; e-mail: pubrights@worldbank.org.

Document and section covers are created using photos by:
Sharath G. / Pexels
Antoni Shkraba/ Pexels
David Kwewum/Pexels
Iqbal Farooz/ Pexels
Abbakar Saeeyd / Pexels
Cottonbro Studio / Pexels
Freepik

ABNASIA.ORG

Acknowledgements

This paper is a product of the Digital Transformation Vice Presidency at the World Bank Group.

It was prepared by Sharmista Appaya (Lead, AI Business Line & Task Team Leader) and Jeremy Ng (Consultant). The team would like to thank the peer reviewers David Leslie (Director of Ethics and Responsible Innovation Research, Turing Institute), David Satola (Lead Counsel), Yan Liu (Senior Economist), Marelize Gorgens (Senior Specialist) and Nay Constantine (Digital Development Specialist), for their constructive comments. Christine Qiang and Peter Kusek provided overall guidance. Thanks to Sajid Chowdhury (Director, Big Blue) for design and layout assistance .

The findings, interpretations, and conclusions expressed in the paper and case studies are entirely those of the authors. They do not necessarily represent the views of the World Bank Group and its affiliated organizations or those of the Executive Directors of the World Bank or the governments they represent.

ABNASIA.ORG

Contents

ABNASIA.ORG

ABNASIA.ORG

Acronyms

AI	Artificial Intelligence
API	Application Programming Interface
CCPA	California Consumer Privacy Act
CDDO	Central Digital and Data Office (UK)
CENELEC	European Committee for Electrotechnical Standardization
CEN	European Committee for Standardization
DFFT	Data Free Flow with Trust
DPI	Digital Public Infrastructure
DSIT	Department for Science Innovation and Technology (UK)
EU	European Union
FDI	Foreign Direct Investment
G7	Group of Seven
GDP	Gross Domestic Product
GenAI	Generative AI
GDPR	General Data Protection Regulation
GPAI	Global Partnership on AI
ICT	Information and Communication Technology
IEEE	Institute of Electrical and Electronics Engineers
ILO	International Labour Organization
IMF	International Monetary Fund
ISO	International Organization for Standardization
ITU	International Telecommunication Union
LGBT	Lesbian, Gay, Bisexual, and Transgender
LGPD	General Data Protection Law (Brazil)
LLM	Large Language Model
MOOC	Massive Open Online Course
NIST	National Institute of Standards and Technology
OECD	Organisation for Economic Co-operation and Development
OSTP	Office of Science and Technology Policy (US)
PAI	Partnership on AI
PDPA	Personal Data Protection Act (Singapore)
PPP	Public-Private Partnership
RAM	Readiness Assessment Methodology
RTA	Responsible Technology Adoption Unit
SDG	Sustainable Development Goal
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
US	United States
WB	World Bank
WBG	World Bank Group
WDR	World Development Report

ABNASIA.ORG

Executive Summary

ABNASIA.ORG

As artificial intelligence (AI) becomes increasingly integral to global economies and societies, the need for effective AI governance has never been more urgent. The rapid advancement in AI technologies, coupled with their widespread adoption across many sectors such as healthcare, finance, agriculture, and public administration, present both unprecedented opportunities and significant risks. Ensuring that AI is developed and deployed in a manner that is ethical, transparent, and accountable requires robust governance frameworks that can keep pace with technological evolution.

This report explores the emerging landscape of AI governance, providing policymakers with an overview of key considerations, challenges, and global approaches to regulating and governing AI. It examines the foundational elements necessary for thriving local AI ecosystems, such as reliable digital infrastructure, a stable and sufficient power supply, supportive policies for digital development, and investment in local talent. As countries navigate this complex landscape, the report highlights the need to encourage innovation by mitigating risks like bias, privacy violations, and lack of transparency, emphasizing the importance of sustainable growth and responsible AI governance.

Regulatory Approaches to AI Governance

The report outlines four key regulatory approaches to AI governance—industry self-governance, soft law, regulatory sandboxes, and hard law—each offering distinct advantages and challenges:

1. **Industry Self-Governance**
   - *Strengths:* Can directly impact AI practices if integrated into business models and company cultures.
   - *Limitations:* Non-binding; not appropriate for sectoral use-cases with particularly high risks – e.g. financial sector or healthcare; risk of 'ethics-washing'.
2. **Soft Law**
   - *Strengths:* Soft law includes non-binding international agreements, national AI principles, and technical standards, providing adaptable frameworks that promote responsible innovation. Early governance efforts by intergovernmental bodies have set important precedents.
   - *Limitations:* While soft law encourages innovation, it focuses on high-level principles rather than binding rights and responsibilities.
3. **Hard Law**
   - *Strengths:* Binding legal frameworks provide clear, enforceable guidelines that ensure AI stakeholders comply with established standards and regulations.
   - *Limitations:* Given the rapid pace of AI development, hard laws risk becoming outdated and can be extremely resource-intensive to implement.
4. **Regulatory Sandboxes**
   - *Strengths:* These controlled environments allow for real-world experimentation with AI technologies, supporting innovation and providing valuable insights without exposing the public to unchecked risks.
   - *Limitations:* Sandboxes can be resource-intensive and have limited scalability, making them less feasible for wide-scale governance across diverse sectors.

Key AI Governance Challenges and Considerations

AI systems are inherently complex and dynamic, with implications that touch on ethical, legal, and socio-economic aspects. Governing AI requires frameworks that promote responsible innovation and risk mitigation, ensuring that AI's benefits are distributed equitably while minimizing potential harms. Moreover, these frameworks must consider sector-specific issues and legacy concerns, particularly in areas like healthcare, finance, and public services, where AI harms can scale rapidly across populations.

One critical challenge is bias and fairness. AI systems, if not properly governed, can perpetuate and even amplify existing societal biases, leading to unfair outcomes,

ABNASIA.ORG

especially in sensitive sectors like criminal justice or healthcare. It is essential that governance mechanisms detect and mitigate bias at every stage of AI development and deployment. Legacy concerns, such as pre-existing societal inequalities, must also be addressed to prevent AI from entrenching or exacerbating these issues.

Another key issue is privacy and security. AI's reliance on vast datasets raises significant concerns about data privacy and security, particularly where sensitive personal information is involved. Robust data protection standards and privacy-preserving AI techniques are necessary to safeguard individual rights and maintain public trust in AI technologies.

Transparency and accountability are equally crucial. AI decisions must be explainable, and developers must be held accountable for the impacts of their systems. Clear standards for explainability, coupled with mechanisms for auditing and oversight, are vital to maintaining public trust. This is especially important in sectors like finance or government, where the stakes are high, and transparency is critical to public confidence.

Lastly, sustainable growth depends on the presence of reliable digital infrastructure, adequate power supply, and a robust talent pipeline. For sectors like agriculture or public administration, where AI can significantly enhance service delivery and efficiency, these foundational elements are crucial. Policymakers must ensure that legacy infrastructure, which may not have been built with AI in mind, is updated to support sustainable and inclusive AI growth.

Key Takeaways

AI governance cannot rely on a single, universal approach, and no regulatory model works in isolation. The report stresses the importance of adopting a flexible, adaptable governance framework that evolves with both technological advancements and societal changes.

Some key takeaways include:

- **Adopting a Multi-Stakeholder Approach:** Policymakers should engage diverse stakeholders—including industry, civil society, and academia—to ensure AI governance frameworks are inclusive, comprehensive, and aligned with ethical standards.
- **Tailoring Regulatory Mechanisms:** Countries must assess the maturity of their AI ecosystem, existing legal and regulatory landscapes, and available resources when determining the most appropriate regulatory mechanisms. A 'one-size-fits-all' approach is unlikely to work given the diversity of AI applications and risks.
- **Promoting International Collaboration:** AI governance is inherently global in scope. As AI technologies transcend borders, international cooperation will be essential to harmonize standards, address cross-border challenges, and ensure AI aligns with global public goods, human rights, and equitable development.
- **Sector-Specific Considerations and Regulatory Legacies:** AI governance frameworks must be tailored to the specific sectors they regulate, recognizing that different industries—such as healthcare, finance, agriculture, and public services —face unique challenges and risks. Additionally, these frameworks must consider the regulatory legacies of individual countries, ensuring that existing legal structures, sector-specific regulations, and data protection laws are integrated into new AI governance models.

The future of AI governance lies in a carefully balanced combination of regulatory mechanisms. Only through this tailored, multi-layered approach can AI's transformative potential be realized for the common good—driving inclusive growth, sustainability, and ethical progress.

Disclaimer

This report is intended to serve as a foundation for broader policy discussions and stakeholder consultations. It does not purport to provide legal, technical, or strategic advice but rather provide an overview of current emerging practices. Please note that issues related to AI adoption, strategic frameworks, and enabling infrastructure are covered in separate papers, which are currently under development.

ABNASIA.ORG

Section 1 Introduction and Background

ABNASIA.ORG

1.1. Introduction

Artificial intelligence (AI) is sparking interest among policymakers globally as a powerful tool to unlock new opportunities for sustainable development. Over 70 countries have already published AI policies or initiatives,[1] with numerous more in progress around the world. AI technology and applications are developing at record pace, as evidenced by the rapid and widespread adoption of generative AI (GenAI) – new tools and applications which create original text, audio, image, and video content. One striking benchmark is to consider the pace at which various technologies have permeated our lives. It took 75 years for fixed telephones to reach 100 million users globally. In contrast, mobile phones achieved this milestone in just 16 years, and the internet took only 7 years. The Apple store took 2 years, and strikingly, ChatGPT reached this number in a mere two months. This unprecedented rate of adoption not only showcases the transformative potential of AI but also sets the stage for a major shift in global connectivity and economic systems.

The responsible adoption of AI has substantial potential to drive inclusive growth and economic development in emerging economies. Investing in AI and digital innovation prepares countries to generate new business models and participate in the global economy. According to UNESCO, AI may add USD $13 trillion to the global economy by 2030 and increase global GDP by 1.2%.[2] It can boost productivity and efficiency in key economic sectors, and in public services to overcome resources gaps, ranging from health and education to transportation and finance. Strategic adoption of technologies can provide employment opportunities for youth, innovators and entrepreneurs to participate in global AI value chains.

AI has the potential to significantly improve efficiency, optimization, and transparency across multiple sectors. For example, cross-sectoral applications such as language translation tools and customer service chatbots can increase access to public services, benefiting both users and administrators. In healthcare, AI tools can help address structural inequalities, shortages of qualified healthcare professionals or supplies, and accessibility barriers when applied responsibly.[3] Similarly, in education, AI can support more inclusive platforms for young children, teenagers, adults, and people with disabilities.[4] In agriculture, AI is used in precision farming, leveraging drone and satellite imagery to support climate adaptation and mitigation outcomes, including forest conservation and the better use of renewable energy. The public sector also benefits from AI, assisting regulators in e.g. the financial sector to support fraud detection and supervision activities.[5]

Although AI has been around for a long time, its impact today is markedly different. The explosion in popularity of GenAI has led to increased focus on AI policymaking and governance. Non-generative algorithmic systems and decision-making processes (referred to as traditional or narrow AI) have been widely used across both the public and private sectors in the past decades. Unlike other emerging technologies such as blockchain, which primarily appealed to niche markets, AI has had significant time to develop and mature, and with a proven track record of enhancing productivity across various sectors.

Policymakers should be aware of the nuances between narrow AI and GenAI, given that the governance interventions might need to be tailored accordingly. GenAI models and large language models (LLMs) are versatile as they are not limited to a specific pre-defined list of 'labels'. This represents a significant shift from narrow AI, traditionally used in

[1] https://oecd.ai/en/dashboards/overview

[2] https://unesdoc.unesco.org/ark:/48223/pf0000382570

[3] https://www.mckinsey.com/industries/healthcare/our-insights/tackling-healthcares-biggest-burdens-with-generative-ai

[4] https://www.oecd.org/en/publications/the-potential-impact-of-artificial-intelligence-on-equity-and-inclusion-in-education_15df715b-en.html

[5] https://www.bankingsupervision.europa.eu/press/interviews/date/2024/html/ssm.in240226~c6f7fc9251.en.html#:~:text=It%20can%20analyse%20vast%20amounts,the%20work%20of%20banking%20supervisors.

ABNASIA.ORG

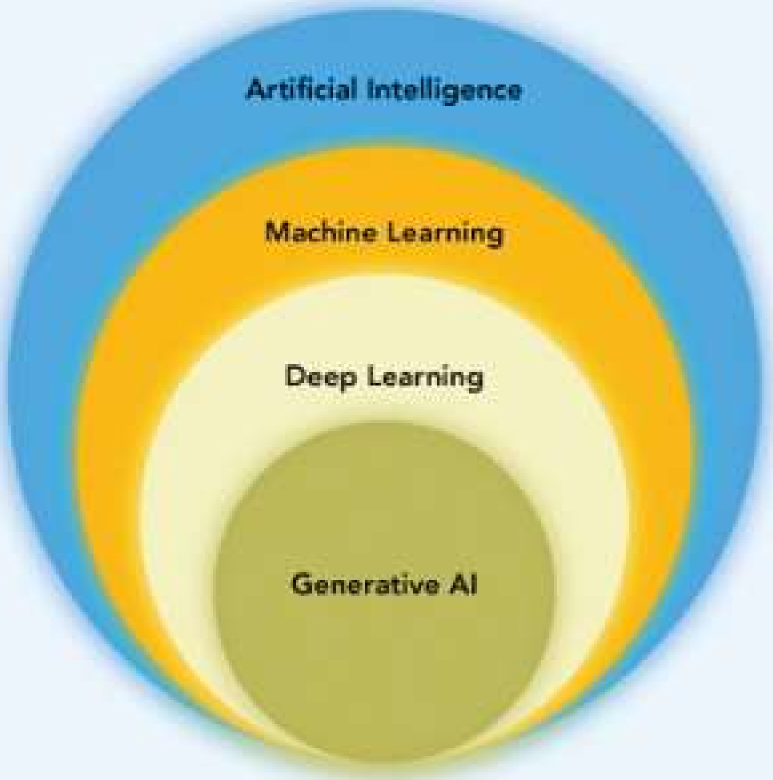


Figure 1. Generative AI fits within the broader context of deep learning, a subset of machine learning.
Source: The World Bank 2024

Box 1: Generative AI: A Primer

Generative AI applications are enabled by large language models (LLMs), which are trained on vast amounts of diverse and unstructured data including original text, audio, images and videos to support the generation of new content. Users can interact with generative AI models through web or application interfaces providing prompts, that the models use to generate original content in various formats.

Leading generative AI models today include OpenAI's GPT series, Anthropic's Claude, Google DeepMind's Gemini and Meta's Llama. These models can generate articles, synthesize text, write poetry, and even create code. They can also respond to questions, engage in discussions, explain complex scientific or social concepts, and provide extensive replies to precise questions and inquiries. Investment and adoption in these generative AI models has been fast paced. For example, OpenAI's ChatGPT application was released in November 2022 and reached 100 million monthly active users in two months, making it the fastest-growing consumer application in history.

Generative AI represents the next frontier in AI, building upon advancements in machine learning, gains in computing power, and the leveraging of extremely large datasets. Whereas a small number of companies are training and developing advanced models, numerous startups, nonprofits, universities, companies and government actors are leveraging existing LLMs to develop their own AI applications. Smaller actors can access pre-trained AI models via their application programming interfaces (API) or model repositories, enabling them to create their own customized applications without the need to train complex models from the ground up. For instance, a startup or government organization might integrate a generative AI model into their applications for purposes such as language translation, customer service, educational content, and more.

Source: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

ABNASIA.ORG

specific applications like image recognition, product recommendation systems, and fraud detection and which depend to a large extent on pre-defined 'labels' on which to train the model. While narrow AI excels within a limited scope, GenAI demonstrates adaptability and competence across diverse contexts (see *Box 2*).

While AI offers numerous benefits, it also presents risks that need to be carefully managed by countries as they engage in the emerging AI economy. AI offers great potential to accelerate productivity, growth, expand economic opportunities, improve societal welfare, and promote inclusion. However, if not managed properly, AI tools and applications also pose significant risks to consumers that, if left unaddressed, could seriously impact fundamental human interests and negatively impact countries' economic growth and development trajectories. The World Bank Digital Progress and Trends Report series outlines several global efforts, concerns and recommendations to address these challenges.[6]

Policymakers should play a proactive role in creating a trust framework for AI governance, promoting adoption of AI by encouraging responsible innovation with proportionate safeguards. This involves establishing

Box 2: Distinguishing Narrow AI and Generative AI

Narrow AI (Traditional AI):

- ***Task-Specific:*** Designed to optimize the efficiency of well-defined, specific tasks.
- ***Pattern Recognition:*** Recognizes features in input data and correlates them with established patterns from training datasets.
- ***Output Type:*** Primarily used in cases where outcomes follow a predictable format, such as generating scores or providing probabilistic classifications.

Generative AI:

- ***Adaptive Learning:*** Learns and adapts from vast and diverse datasets.
- ***Content Creation:*** Capable of producing original content including text, audio, images, and videos based on input prompts.
- ***Output Type:*** Used in creative and dynamic tasks, such as generating articles, code, images, and engaging in complex discussions. More likely to succeed with unstructured imagery or natural language interfaces.

Key Differences:

- ***Learning Approach:*** Narrow AI often requires labeled training data, while Generative AI learns from larger sets of unstructured data.
- ***Output Nature:*** Narrow AI tends to provide responses from a set range of options, whereas Generative AI can generate dynamic, language-based, or visual responses.
- ***Flexibility:*** Generative AI can handle a wider variety of tasks and adapt to new data more fluidly than Narrow AI.
- ***Infrastructural Requirements (compute and data):*** Traditional AI approaches can usually trace and evaluate the appropriateness of their training data and produces algorithms with relatively low computational cost. Generative AI algorithms require vast amounts of data and require substantial compute resources for both model training and inference.

*Adapted from inputs from multiple sources.

[6] https://www.worldbank.org/en/publication/digital-progress-and-trends-report?cid=ECR_LI_worldbank_EN_EXT

ABNASIA.ORG

comprehensive policy and regulatory frameworks, building the enabling foundations for AI innovation and ecosystems to thrive, and addressing human capital needs and access to digital infrastructure, computing resources, and datasets. Targeted policies can support AI adoption in key sectors and foster the growth of local innovation ecosystems. Additionally, AI harms can be managed through a combination of regulatory approaches, including binding laws and regulations, technical standards, international and national ethical principles, and private codes of practice.

Looking ahead, international standards-setting and cooperation are important to guide responsible AI adoption for sustainable, inclusive, and resilient growth. These principles are illustrated with country examples throughout the report, showcasing developing strategies and practices in AI governance and regulation.

This report seeks to provide policymakers with an overview of current approaches to creating robust, fit-for-purpose national AI governance[7] frameworks. To meet fast-changing technological and societal trends, agile and flexible policymaking is essential. Multi-stakeholder participation, especially consultation with consumers and affected communities, along with international and regional coordination, are crucial in the design and implementation of AI governance and policy frameworks. As AI development and deployment advance, policymakers must be informed, coordinated, and equipped to respond to both new opportunities and disruptions. Regulation, where needed, must be technology-agnostic, focusing on outcomes and principles. Societal and cyber resilience, AI and digital literacy and inclusion, and sustainability are also important considerations.

Section II highlights the foundational elements needed to create an enabling environment for AI; Section III outlines the promises and challenges of AI and the difficulties in regulating it; Section IV then examines various regulatory tools and highlights some key principles for policymakers to consider as they design their approach to AI governance. Section V sets out key dimensions for AI governance, while Section VI outlines the stakeholder ecosystem and common institutional arrangements for oversight of AI; finally, Section VII looks to the future, offering some parameters and recommendations for policymakers as they develop their AI governance frameworks.

The report surveys different types of AI governance arrangements around the world illustrated through country examples. Although it is too early to definitively say what has worked best, these principles highlight developing strategies and practices in AI governance and regulation. Reliable digital infrastructure, sufficient and stable power supply, policies enabling digital development and investment in local talent are some of the foundational requirements for local AI ecosystems. This section sets out essential prerequisites that can act as enabling foundations for countries seeking to harness the benefits of AI for sustainable development.

[7] In this report, the term 'governance' refers to the broader framework of laws, rules, practices, and processes used to ensure AI technologies are developed and used responsibly. The term 'regulation' is used in a narrower sense to refer to binding legal or regulatory guardrails imposed on AI developers and deployers.

ABNASIA.ORG

Section 2 Enabling Foundations for AI

ABNASIA.ORG

Reliable digital infrastructure, sufficient and stable power supply, policies enabling digital development and investment in local talent and are some of the foundational requirements for local AI ecosystems. This section sets out essential prerequisites that can act as enabling foundations for countries seeking to harness the benefits of AI for sustainable development.

2.1. Digital and data infrastructure

The successful deployment of AI technologies in a country hinges on robust digital and data infrastructure. This foundation is essential to support the development, deployment, and scaling of AI applications across various sectors. Key components of this infrastructure include high-speed internet, data storage and management systems, and computational power.

a. High-Speed Internet

High-speed internet is the backbone of digital infrastructure. It ensures that data can be transmitted quickly and efficiently between devices, data centers, and cloud services. For instance, countries like South Korea and Singapore have achieved internet speeds exceeding 200 Mbps, enabling seamless AI operations and real-time data processing. In contrast, countries with slower internet speeds face significant delays in data transmission, hindering AI application performance.

b. Devices

The availability of devices such as computers, smartphones, and IoT devices plays a crucial role in the development, deployment, and utilization of AI technologies. Devices like smartphones, computers, and IoT devices gather vast amounts of data essential for training AI models and enable real-time processing through edge computing, reducing latency and enhancing privacy. They also democratize AI by making it accessible to a broader population, allowing more people and organizations to develop and benefit from AI technologies. However, the share of mobile phone owners is only 49 percent[8] in low-income countries. This lack of access hinders inclusive AI growth and constrains the collection of diverse and representative data crucial for developing relevant AI algorithms.

c. Data Storage and Management Systems

AI applications generate and rely on vast amounts of data. Efficient data storage solutions, such as data lakes and warehouses, are critical to managing this data and training AI models. Moreover, proper data management systems ensure data integrity, accessibility, and security. According to a report by Gartner, global spending on data storage is expected to reach $25 billion by 2025, reflecting the growing importance of this infrastructure component.

d. Computational Power

Compute capacity—the ability to store, process, and transfer data at scale[9]—is crucial for training and deploying AI models and applications. High-performance computing (HPC) and Graphics Processing Units (GPUs) are pivotal in this context. Affordable access to international cloud computing services is a valuable resource for both training—teaching a model to recognize patterns in data—and AI inference, applying the trained model to new data to generate predictions or decisions. Training requires significantly more computational power than inference.

For example, training OpenAI's GPT-3 involved processing 570 gigabytes of text data using thousands of GPUs over several weeks, whereas inference tasks using GPT-3 require approximately 1-2 orders of magnitude less compute power, often only needing a single GPU or a small cluster of GPUs for real-time processing.[10] However, a number of countries, face challenges in scaling their computational infrastructure. The reliance on international cloud computing services can be expensive and may not always meet the specific[11] needs of local AI practitioners. Furthermore, dependence

[8] https://www.worldbank.org/en/publication/digital-progress-and-trends-report?cid=ECR_LI_worldbank_EN_EXT

[9] Definition of compute by Tony Blair Institute for Global Change. Retrieved from https://www.institute.global/insights/tech-and-digitalisation/state-of-compute-access-how-to-bridge-the-new-digital-divide

[10] https://www.springboard.com/blog/data-science/machine-learning-gpt-3-open-ai/

[11] https://indiaai.gov.in/

ABNASIA.ORG

on external providers can pose risks related to data sovereignty, privacy, vendor lock-in and security. Investments in HPC and local cloud infrastructure are crucial for fostering a sustainable and competitive AI ecosystem.

Regional collaborations can consolidate resources towards shared data centers. For example, the European High Performance Computing Joint Undertaking (EuroHPC JU)[12] is a significant initiative aimed at pooling resources across European countries to develop a world-class supercomputing ecosystem. Moreover, as demand for edge computing grows, investments in local infrastructure become even more critical. Edge computing reduces latency by processing data closer to where it is generated, which is particularly important for applications requiring real-time processing and decision-making.[13]

It should be noted however that mitigating the environmental impacts of AI is also an important consideration. Evidence shows sharply increased water and electricity consumption due to AI training and development. Developing more energy-efficient algorithms and sustainable AI infrastructure powered by clean energy is crucial for addressing these challenges and ensuring long-term sustainability and competitiveness.[14]

e. High Quality Multimodal Data

High-quality multimodal data is the backbone of the digital economy and a crucial element for AI development. This type of data encompasses various formats, including text, images, audio, and video, allowing AI models to understand and process information from multiple sources effectively. For example, combining textual data with visual and audio data can enhance an AI system's ability to recognize speech, understand context, and make accurate predictions.

However, disparities in digital access lead to underrepresentation within datasets, resulting in less representative training data. This, in turn, lowers the accuracy of model outputs and can potentially cause biased or harmful outcomes. The issue is further exacerbated when AI models are trained on foreign datasets that are not suited to local contexts, leading to inaccurate, unsafe, and discriminatory outcomes.

To address these challenges, governments could increase the availability of AI-ready open datasets by digitizing local data, including public sector records, and making it publicly accessible. Synthetic data has also been explored to enhance datasets, but it should be carefully managed to avoid perpetuating biases.[15] Public and private sector entities can then utilize these open datasets to develop consumer-beneficial products.

Countries that invest in collecting and curating diverse, high-quality multimodal datasets are better positioned to develop advanced AI applications that are more accurate, reliable, and capable of performing complex tasks across different domains. However, expanding data access must be balanced with good data governance and sharing practices, ensuring privacy, security, and fair representation to support trustworthy and inclusive AI systems. There is an urgent need for critical research on the intersection of data and AI governance. For example, the need to combat algorithmic bias in outputs by using larger, more representative, and inclusive datasets to train AI models, may sit in tension with core data protection principles such as data minimization.

[12] https://eurohpc-ju.europa.eu/about/discover-eurohpc-ju%5Fen

[13] The State of AI Infrastructure at Scale 2024

[14] Gartner. (2023). 'Market Guide for Cloud Infrastructure as a Service.' Retrieved from Gartner.

[15] Datasets require sufficient real data in each generation to ensure their quality (precision) or diversity (recall). https://arxiv.org/abs/2307.01850

ABNASIA.ORG

Box 3: Country Example: India

India's comprehensive AI Mission recognizes the critical importance of computational power as a prerequisite for AI development. With a strong emphasis on building and democratizing computational infrastructure, the mission has a budget outlay of Rs.10,371.92 crore. (USD 1.38 billion). Beyond computational power, the AI mission encompasses several other key components designed to foster innovation, ensure ethical practices, and drive socio-economic transformation.

Key Components of the IndiaAI Mission:

- *High-End Scalable AI Computing Ecosystem:*

  The mission includes the establishment of a high-end scalable AI computing ecosystem with over 10,000 Graphics Processing Units (GPUs), built through public-private partnerships. This infrastructure is designed to meet the demands of India's rapidly expanding AI start-ups and research ecosystem.

- *AI Marketplace:*

  An AI marketplace will be developed to offer AI as a service and provide pre-trained models to AI innovators. This marketplace will serve as a one-stop solution for critical AI resources, facilitating easy access and promoting innovation.

- *IndiaAI Innovation Centre:*

  This centre will focus on the development and deployment of indigenous Large Multimodal Models (LMMs) and domain-specific foundational models across key sectors. It aims to bolster India's capabilities in AI and ensure the development of AI solutions that cater to local needs.

- *IndiaAI Datasets Platform:*

  A unified platform will be created to streamline access to quality non-personal datasets, ensuring that Indian startups and researchers have seamless access to the data necessary for AI innovation.

- *IndiaAI Application Development Initiative:*

  This initiative will promote AI applications in critical sectors by developing, scaling, and promoting impactful AI solutions with the potential for large-scale socio-economic transformation.

- *IndiaAI FutureSkills:*

  The program aims to mitigate barriers to AI education by increasing the availability of AI courses at undergraduate, masters, and Ph.D. levels. Data and AI labs will be established in Tier 2 and Tier 3 cities to offer foundational AI courses, ensuring that AI education is accessible across the country.

- *IndiaAI Startup Financing:*

  This pillar will support and accelerate deep-tech AI startups by providing streamlined access to funding, enabling them to undertake futuristic AI projects and drive innovation.

The IndiaAI mission is poised to create highly skilled employment opportunities, leverage the country's demographic dividend, and enhancing India's global competitiveness.

Source: https://indiaai.gov.in/

ABNASIA.ORG

Box 4: Korea's Data Dam Initiative

South Korea has launched an ambitious project known as the Data Dam, aimed at enhancing the country's data infrastructure and fostering innovation in AI and big data. This initiative is part of the Korean New Deal, which focuses on digital transformation and green growth.

The Data Dam project involves collecting and utilizing vast amounts of data across various sectors, including healthcare, transportation, and finance. By integrating data from multiple sources and making it accessible through a centralized platform, Korea aims to create a robust data ecosystem that supports the development of AI applications. Just like a water-storage dam collects, stores, and distributes water to the surrounding land for activities such as farming, the Data Dam project collects information from public and private sectors to create useful data and releases it across all industries.

Key features of the Data Dam initiative include:

- *Centralized Data Integration:* Combining data from public and private sectors into a unified platform to break down silos and promote efficient data use.
- *AI Hub Establishment:* Creating an AI hub to provide companies and researchers with access to AI training data from the Data Dam and cloud-based high-performance computing resources.
- *Sectoral Data Utilization:* Focusing on sectors such as healthcare, transportation, and finance to drive innovation and improve services through AI applications.
- *Data Privacy and Security:* Implementing robust data protection measures to safeguard personal information and comply with regulations, thus building public trust.

The Data Dam initiative has already shown promising results, with significant progress in data collection, integration, and utilization.

Source: Ministry of Science and ICT, South Korea. 'Korean New Deal'; Korea Data Agency. 'Data Dam Initiative.'; OECD

2.2. Human Capital (AI and Digital Readiness)

Governments must adapt education and training programs to prepare workforces for participation in the global AI value chain while mitigating labor market disruptions and potential job losses due to automation.[16] The AI value chain offers employment opportunities across skill levels, from data collection and preparation to machine learning research and management of data centers and cloud infrastructure. While some outsourced jobs may face automation, countries can target AI adoption towards technologies that leverage labor and address domestic needs.

This effort should include both upskilling current workers to enhance their existing capabilities and reskilling individuals to equip them with new skills for emerging job opportunities. Additionally, it is crucial to focus on capacity building within government institutions to ensure they have the expertise required to effectively regulate and govern AI technologies.

[16] Value chains are sequences of processes involved in the creation, development, deployment, and utilization of AI technologies and solutions. Including data collection and processing, algorithm design and development, model training and optimization and integration among others. Subject of forthcoming WB paper.

ABNASIA.ORG

Investing in productivity, and digital skills will be key to augmenting the labor force with AI rather than replacing it.[17] Education and training programs should equip students with skills in machine learning, data science, business, data engineering, computer science, and practical technical skills like data center maintenance or data preparation and management.[18] Most regions lack human capital and a talent pool ready to develop or apply AI applications. Networks of exchange among university professors, such as those established by the African Institute for Mathematical Sciences,[19] can help overcome shortages in knowledgeable lecturers at low cost. Training programs must emphasize inclusivity, particularly targeting rural communities and women, to prevent widening inequality and divides.

Moreover, there is a need for capacity building within government institutions to ensure they have the expertise required to effectively regulate and govern AI technologies. Education and training programs should also consider fostering so-called 'soft skills', those that AI cannot easily replicate, such as judgment, critical thinking, and emotional intelligence. Measures to improve digital literacy are important, as it remains a significant hurdle to the development, management, adoption, and use of AI, particularly in low-income countries (LICs). Addressing this foundational challenge is crucial for enabling the wider population to participate in and benefit from AI-driven economic opportunities.

2.3. Local Ecosystem

This section does not go into the details of the enabling ecosystem but is here to illustrate its importance as a foundational element for AI development.

A robust ecosystem is essential for fostering AI development and adoption, complementing digital and data infrastructure and human capital. This ecosystem includes elements such as **research and development (R&D)** - crucial for advancing AI technologies and can potentially include government funding, private sector investment, and academic partnerships; **public-private partnerships (PPP)**- which can help in pooling resources, sharing knowledge, and driving large-scale AI projects; a vibrant **startup ecosystem** including support for startups, access to funding, shared infrastructures such as incubators and accelerators, and the tools to support collaborations among industry players, academics, local organizations, and other community stakeholders; and **awareness and advocacy** about AI and its potential benefits to drive adoption both in the public and private sector.

Governments can lead by example, through promoting **internal AI adoption** or offering subsidies to solve pressing challenges in key industry sectors such as healthcare, education, environment, energy or beyond. Some examples of this include India's 'AI for All' approach, a self-learning online program designed to raise public awareness of AI for inclusive development, highlighting AI startups addressing social challenges in healthcare, language translation and agriculture.[20] Additionally, governments can create an enabling environment for AI investment through supportive policies, seed investment funds and co-financing, incentives or even support public procurement by pre-certifying certain AI vendors, such as in Canada's List of AI Suppliers, to facilitate the integration and adoption of AI technologies.[21]

More details on the ecosystem can be found in our forthcoming toolkit on developing a country-specific AI Strategy.

[17] Digital Progress and Trends Report 2023, The World Bank, 2024, https://openknowledge.worldbank.org/server/api/core/bitstreams/95fe55e9-f110-4ba8-933f-e65572e05395/content

[18] https://www.oecd.org/publications/the-impact-of-ai-on-the-workplace-main-findings-from-the-oecd-ai-surveys-of-employers-and-workers-ea0a0fef-en.htm and https://www.oecd.org/els/the-impact-of-ai-on-the-workplace-evidence-from-oecd-case-studies-of-ai-implementation-2247ce58-en.htm

[19] https://nexteinstein.org/

[20] AI for All, India

[21] https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/list-interested-artificial-intelligence-ai-suppliers.html

ABNASIA.ORG

Box 5: Singapore's AI Apprenticeship Program

Singapore's AI Apprenticeship Program (AIAP) has successfully trained over 300 Singaporeans, equipping them with practical AI technical skills to meet the growing demands of the domestic AI ecosystem. This full-time program runs for 9 months and is structured into two phases: a 2-month intensive deep-skilling training followed by a 7-month real-world AI project. During the program, apprentices are paired with mentors and gain access to industry recruitment opportunities.

The AIAP is fully funded by the government and includes a monthly stipend for apprentices, which varies based on their years of relevant work experience and qualifications. The program is inclusive, welcoming participants of various ages, with special provisions for Singaporeans aged 40 years or above who are eligible for an extension to gain additional business and technical hands-on experience.

To be eligible for AIAP, applicants must be Singaporean citizens, graduates from a recognized university or polytechnic, and possess prerequisite programming competencies. AIAP is part of the national AI Singapore initiative, supported by the National Research Foundation and hosted by the National University of Singapore

Source: https://aisingapore.org/aiap/ ; https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2023/imda-leads-ai-skilling-to-build-ai-talent-pool

ABNASIA.ORG

Section 3 The Promise and Perils of AI

ABNASIA.ORG

Despite the transformative potential of AI across multiple sectors, there are important practical challenges to implementation – crucially, robust governance frameworks are needed to ensure AI systems are trusted by consumers.

AI systems present several existing risks that stem from their inherent limitations and the quality of the data they are trained on. One of the most prominent risks is bias and discrimination. AI models can perpetuate and even exacerbate existing biases if they are trained on unrepresentative or biased datasets. This can lead to unfair treatment and outcomes, particularly for underrepresented and marginalized groups. For example, facial recognition systems have been shown to have higher error rates for people with darker skin tones compared to those with lighter skin tones[22], raising serious concerns about their use in law enforcement and surveillance. Additionally, the lack of explainability and transparency in AI decision-making processes makes it difficult to identify, audit, and rectify these biases, further compounding the risk of discrimination. As AI systems play an increasingly significant role in decision-making processes across sectors, the lack of explainability remains a barrier to auditing and improving the models.

Moreover, while AI is being applied in use cases for environmental and climate protection, such as predicting and monitoring deforestation patterns or optimizing renewable energy systems, **the AI model supply chain consumes a huge amount of energy, water, and other natural resources.** This contributes to increased carbon emissions and potential environmental degradation, highlighting the need for sustainable practices in AI deployment. For example, LLMs emit up to 550 tons of CO2 during their training processes. Serious sustainability concerns also apply to model inference processes – Google attributes 60% of its AI-related energy use to inference;[23] generating one image using AI uses the same amount of energy as charging a smartphone.[24] Large tech companies are at risk of missing their climate targets – with Microsoft and Google both announcing in 2024 that they would miss their sustainability targets set during previous years.[25] There are also increasing concerns regarding the water consumption needed to cool the computing equipment housed within data centers – Microsoft has noted that 42% of the water it consumed in 2023 came from 'areas with water stress'.[26] Addressing these challenges requires rethinking the dominant 'bigger is better' paradigm and deepening appreciation of the value of smaller AI models, mandating greater transparency in terms of compute cost and energy usage, and promoting research that focuses on resource efficiency[27] – this requires collaborative efforts across sectors and borders, robust policy frameworks, and ongoing research and development to ensure that AI technologies are implemented responsibly and equitably.

Beyond well-documented risks such as biases and lack of explainability, newer challenges are emerging as AI technologies evolve. One such risk associated with GenAI in particular is the phenomenon of AI hallucinations, where AI systems generate outputs that are factually incorrect yet appear plausible. The complexity and opacity of these models make it difficult to predict and control when hallucinations will occur, posing significant risks in critical applications such as healthcare, legal advice, and education.

[22] Buolamwini, J., & Gebru, T. (2018). 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.' Proceedings of the 1st Conference on Fairness, Accountability and Transparency. PMLR 81:77-91. Available at: http://proceedings.mlr.press/v81/buolamwini18a.html

[23] https://arxiv.org/pdf/2409.14160

[24] https://arxiv.org/pdf/2311.16863

[25] https://arxiv.org/pdf/2409.14160

[26] https://techcrunch.com/2024/08/19/demand-for-ai-is-driving-data-center-water-consumption-sky-high/

[27] https://arxiv.org/pdf/2409.14160#page=13&zoom=100,48,86

Box 6: AI Risks

1. **Bias and Discrimination:** AI systems can perpetuate bias and discrimination due to unrepresentative datasets and a lack of transparency in algorithms.

2. **Labor Market Disruption:** The adoption of AI technologies can lead to significant labor market disruption, resulting in job losses and a widening digital divide.

3. **Misuse of AI & Trust Erosion:** AI can be misused for spreading misinformation, creating deepfakes, conducting cybercrime, interfering with elections, and facilitating fraud and scams, which erodes trust in public and private institutions.

4. **Inequality and Access:** There are growing gaps in inclusion and widening inequality based on differential access to AI technologies.

5. **Environmental Impacts:** AI systems, particularly those involving large-scale data processing and machine learning models, consume significant amounts of energy, contributing to environmental degradation and increased carbon emissions.

6. **Cybersecurity Vulnerabilities:** AI systems and applications are susceptible to various cybersecurity vulnerabilities due to their complexity and multiple points of vulnerability. LLMs and other foundation models currently lack adequate security requirements.[28] Critical services and infrastructure may become inaccessible due to AI failures or targeted cyber-attacks.

7. **Privacy and Data Protection:** AI-driven surveillance and misuse of personal information pose significant privacy risks. Training AI models requires huge amounts of data, leading to significant concerns regarding mass data collection and processing of personal data.

8. **Physical Safety Risks:** Additionally, AI system failures, security breaches, or unintended AI behavior.

9. **Explainability and Accountability:** The lack of explainability and accountability in AI decision-making processes raises serious concerns, especially where end-users wish to challenge certain algorithmic decisions.

10. **Risks related to deployment context:** Depending on the context in which the AI system is deployed, a range of risks can arise. For example, if not deployed appropriately, generative AI tools used by students may threaten learning quality by lowering retention due to a deepened dependency of students on AI tools. In healthcare contexts, AI systems used for disease prediction and diagnosis that are not robustly designed may lead to biased results, under- or mis-diagnosis, and potentially delayed treatment.

11. **Geopolitical Risks:** The development and deployment of AI in certain sectors can lead to geopolitical instability, e.g. by increasing fragility and conflict via the use of autonomous weapons.

12. **Social and Cultural Impact:** The integration of AI can disrupt social norms and lead to cultural homogenization.

13. **Intellectual Property:** Mass data collection raises concerns regarding the legality of using copyrighted material and other information protected by IP law to train AI models.

14. **Psychological Impact:** The influence of AI on mental health and human-AI interaction dynamics can have profound psychological effects.

Source: adapted and updated from https://openknowledge.worldbank.org/server/api/core/bitstreams/9040dbbb-8594-4083-a399-24192313f907/content

Disclaimer: Non-exhaustive

[28] https://www.rand.org/pubs/working_papers/WRA2849-1.html

ABNASIA.ORG

Box 7: Bias in AI system leads to exclusion of families from childcare benefits in the Netherlands

An AI system employed by the Dutch tax authority inaccurately excluded eligible recipients from welfare benefits, causing significant negative repercussions. The Dutch tax authorities employed an AI tool to create risk profiles for identifying child care benefits fraud. However, this system inaccurately labeled tens of thousands of families, often lower-income or ethnic minorities, as fraudsters based on flawed risk indicators like having dual nationality or low income. As a result, many families faced severe consequences, including crippling debts to the tax agency which pushed them into poverty, loss of child custody, and in some cases, suicide. More than one thousand children were taken into foster care. This incident underscores how AI bias and automation can lead to the inaccurate exclusion of vulnerable populations from important public assistance.[29] It also highlights the need for robust regulations, algorithmic transparency, human oversight, and avenues for redress when automated decisions cause harm.

Source: https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/

Additionally, AI poses risks from job automation and labor market disruptions. The International Monetary Fund (IMF) estimates that nearly 40% of jobs in emerging markets and 26% in low-income countries are exposed to AI, compared to 60% in advanced economies due to the prevalence of cognitive tasks. While emerging markets and developing countries (EMDEs) are less exposed to job disruptions, they are also less equipped to benefit from productivity gains, exacerbating the digital divide and income disparity within and among countries.[30] While the forthcoming effects of AI on labor markets are still unknown, there is notable potential for job losses, increased inequality, and societal disruptions. Addressing these new and evolving risks requires ongoing research, robust verification mechanisms, and stringent oversight to ensure AI systems are reliable and trustworthy.

[29] https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/

[30] https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2024/01/14/Gen-AI-Artificial-Intelligence-and-the-Future-of-Work-542379

ABNASIA.ORG

Box 8: Understanding AI Hallucinations

AI Hallucinations occur where AI systems, particularly those powered by LLMs, produce outputs that are plausible sounding but factually incorrect or nonsensical. These errors occur because the AI generates text based on patterns and data it has been trained on, without an understanding of the real-world context or factual accuracy.

Example of an AI Hallucination

Consider an AI chatbot designed to assist with medical queries. A user might ask, 'What are the symptoms of a heart attack?' An accurate response would include symptoms such as chest pain, shortness of breath, and dizziness. However, an AI hallucination might generate an answer like, 'Heart attacks can be treated effectively with green tea and meditation,' which is misleading and potentially dangerous.

Real-World Instance

In 2020, OpenAI's GPT-3 was noted for generating a response suggesting that 'Ebola is caused by spirits.' This statement is a clear hallucination, as Ebola is a viral infection caused by the Ebola virus, and the response lacks any scientific basis. Such instances highlight the critical need for verifying AI-generated information, especially in sensitive domains like healthcare.

Mitigating Hallucinations

To reduce the risk of AI hallucinations, it is crucial to:

- Implement robust verification mechanisms to check the factual accuracy of AI outputs.
- Use domain-specific training data to improve the contextual accuracy of AI models.
- Continuously monitor and update AI systems to correct and learn from mistakes.
- Ensure users of GenAI systems are correctly trained to identify and manage hallucinations.

Source: Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?' Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. ACM Digital Library; Marcus, G., & Davis, E. (2020). 'GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about.' MIT Technology Review. MIT
Technology Review.

ABNASIA.ORG

Box 9: Open-source AI amplifies benefits and risks

'Open-source' AI models are those whose source code is openly shared under a licensing model that grants users the right to access, modify, and redistribute code. The term is often associated with AI models that have widely available and publicly accessible components such as model weights, training data or code. Several advanced AI models, including LLMs developed by major tech companies, are available under open-source licenses, enabling local AI practitioners to use them without paying licensing fees for context-specific projects while also expanding the potential for misuse.[31]

The open-sourcing of AI models or public accessibility of model components can 'democratize' access to AI by allowing more actors to adapt models for local context, provided they have the necessary infrastructure, data, and skills. It also enables greater transparency, allowing external parties to conduct inspections, audits, research, and bug detection.

Conversely, open-source models can be more easily misused. Access to model weights can compromise the safety of models by allowing actors to remove safety guardrails, potentially generating harmful outputs.[32] As AI systems become more capable, the potential for misuse and harm grows, and practitioners may lack the tools or awareness to apply models responsibly. Once LLM model weights are made public, it is infeasible to monitor, retract, or stop their use, as models may be copied and distributed.[33] While many open-source projects use licenses that promote responsible use to a limited degree, governments and societies should anticipate and prepare for harms and misuse in the absence of comprehensive safeguards or global regulation.

Source: https://spectrum.ieee.org/open-source-ai-2666932122; BadLlama: cheaply removing safety fine-tuning from Llama 2-Chat 13B, https://arxiv.org/pdf/2311.00117.pdf. See page 3, 'for open-source models safety filters can simply be removed by deleting a few lines of code.' https://arxiv.org/pdf/2211.14946.pdf

[31] https://spectrum.ieee.org/open-source-ai-2666932122

[32] BadLlama: cheaply removing safety fine-tuning from Llama 2-Chat 13B, https://arxiv.org/pdf/2311.00117.pdf. See also page 3. 'for open-source models safety filters can simply be removed by deleting a few lines of code.' https://arxiv.org/pdf/2211.14946.pdf

[33] https://spectrum.ieee.org/open-source-ai-2666932122

ABNASIA.ORG

3.1. Challenges in Governing AI

There are various challenges involved in governing AI. Some of the most pertinent include:

1. **Keeping pace with technological advancements.** One of the primary issues is the rapid pace of AI development. As highlighted by Stanford University's AI Index Report 2024, investment in generative AI accelerated to $25.2 billion in 2023, with applications spanning customer support, healthcare, autonomous vehicles, fintech, drones, legal tech, and manufacturing.[34] This rapid evolution, often referred to as the 'pacing problem,' means that regulatory and governance frameworks struggle to keep up. Developing new laws and policies can take months or even years, during which AI technologies continue to advance, creating governance gaps.

2. **Limited technical expertise and knowledge gaps.** Another challenge is limited technical expertise within governments. Policymaking is hindered by knowledge gaps regarding AI technologies and their applications. Higher salaries in the private sector contribute to a brain drain, with a significant proportion of AI talent opting for private or international roles over government positions. For instance, only 0.7% of new AI PhD graduates in the United States and Canada choose to work in government roles.[35] This lack of expertise makes it difficult to draft effective policy, regulatory, and governance measures.

3. **Sector-Specific Governance Needs:** AI governance needs to be tailored to different sectors, each with unique requirements, risks, and operational contexts. For example, the healthcare sector prioritizes patient privacy and safety, necessitating stringent regulations to protect sensitive health data and ensure the reliability of AI-driven diagnostic tools. Conversely, the financial sector focuses on fraud detection, risk management, and compliance with financial regulations. Similarly, the transportation sector must address safety and efficiency in AI applications for autonomous vehicles and traffic management. Powerful foundation models present unique governance challenges due to their broad applicability across various sectors. These models require comprehensive governance measures that go beyond traditional sector-specific approaches, necessitating coordination across multiple government entities and sectors. These sector-specific differences highlight the importance of developing customized governance frameworks to ensure responsible and effective AI implementation across diverse domains.

4. **Cross-Jurisdictional Coordination:** AI development, deployment, and use are often cross-jurisdictional, necessitating international coordination. AI models may be trained on datasets collected from numerous countries and accessed through international cloud services. Different stages of AI development occur in multiple jurisdictions with varying legal frameworks, making it challenging for individual countries to regulate the entire AI lifecycle. The material AI supply chain involves materials, hardware, and labor sourced from a wide array of countries across both the Global North and South. Without a coordinated global approach, disparate national policies can lead to regulatory arbitrage, inconsistencies, and potential loopholes, resulting in gaps or even a 'race to the bottom' in AI governance.

5. **Complexity of AI Supply Chains:** The complex supply chains of AI products, particularly generative AI and LLMs, present significant challenges for governance and accountability. These AI systems often rely

[34] https://aiindex.stanford.edu/report/; Figure 4.3.3 page and Figure 4.3.15 page 254.

[35] https://aiindex.stanford.edu/report/; figure 6.1.7 'Employment of new AI PhDs (% of total) in the United States and Canada by sector, 2010-22, page 335.

ABNASIA.ORG

on vast amounts of data sourced from multiple providers and specialized hardware and software components supplied by different vendors. The complexity and lack of transparency in these supply chains make it difficult to trace the origins of potential issues and identify practical points for regulatory intervention.

6. **Balancing Innovation and Risk Mitigation:** Ensuring governance approaches take a proportionate approach to promoting AI innovation while mitigating potential risks is a delicate task. Disproportionate regulatory provisions can over-burden startups with limited compliance resources, while insufficient governance leaves individuals and society vulnerable to serious risks. Governing AI involves addressing complex ethical, technical, and socio-economic challenges, hence policymakers must create adaptable governance frameworks that provide clear guidelines and safeguards that enable rather than hinder responsible technological progress.

Box 10: **When should AI governance policies be introduced?**

One of the core challenges of AI governance is correctly timing policy interventions. Early on in the AI adoption lifecycle, we face an 'information' problem: given the rapid pace of cutting-edge AI development, it is difficult to predict how AI's critical features, uses, and risks will evolve over time. However, if AI adoption becomes widespread, policymakers may face a 'control' problem: exercising governance control over AI systems may become harder because AI approaches, applications and structures become entrenched in path-dependent ways.

Given the nature of this dilemma, it is impossible to set out a prescriptive, one-size-fits-all approach – however, two key principles may help policymakers navigate these issues for their local contexts. First, thinking of governance as an iterative, agile process can enable policy interventions to be tailored and updated as technology develops and new information is collected. Second, collaborative multi-stakeholder approaches to governance can increase the degree of openness and transparency regarding how governance decisions are made – enabling greater trust between all societal stakeholders and the technologies being governed.

Source: https://demoshelsinki.fi/2022/02/15/what-is-the-collingridge-dilemma-tech-policy/

ABNASIA.ORG

Section 4

Regulatory and Policy Frameworks

ABNASIA.ORG

Policymakers seeking to craft robust AI governance frameworks are faced with several complex challenges. On one hand, there is an urgent need to establish robust governance frameworks to ensure the ethical, fair, and responsible use of AI technologies. Without such frameworks, there is a risk of AI systems perpetuating biases, infringing on privacy, and making decisions without accountability. On the other hand, technology-specific governance interventions may provide clearer guidelines tailored to the unique challenges of AI, but they risk becoming quickly outdated due to the rapid pace of technological advancement. Conversely, tech-agnostic interventions, which focus on broader principles applicable across various technologies, offer flexibility and longevity but may lack the specificity needed to address AI's unique risks and opportunities. Striking the right balance between these approaches is critical to fostering innovation while safeguarding societal values and human rights.

For the potential benefits of AI to be realized, all societal stakeholders must trust the AI systems and institutions that they are engaging with. The UN High-Level Advisory Body on AI has noted that governance is a key enabler and precursor for responsible AI.[36] They indicate that the creation of AI systems that work towards the Sustainable Development Goals (SDGs) cannot be guided solely through market forces or self-regulation by the private sector; they require concerted governmental and intergovernmental policymaking and coordination.[37] Building on the approach set out in the World Bank's World Development Report 2021: Data for Better Lives, this paper sets out the different legal, regulatory and governance tools available for creating trust in the AI ecosystem, encompassing both safeguards - to prevent AI harms and enablers - to facilitate and encourage responsible AI innovation) .[38] Where possible this has been illustrated with country examples.

Some policymakers may be concerned about the risk of over-regulating nascent AI industries and stifling innovation. As such, it is important to ensure any regulatory interventions are proportionate and tailored to the risks, harms and potential societal impact of the AI systems being regulated. At the same time, **policymakers should note that the empirical relationship between regulation and innovation is highly unclear;[39] often, regulation is critical for creating a level competitive playing field for new market entrants while also creating legal certainty for established AI developers and deployers.** Clear regulations help companies plan and invest with confidence, knowing the standards they must meet. Conversely, leaving AI systems unregulated risks exposing consumers to unacceptable harms and leaves critical decisions regarding AI deployment to market forces and private companies, potentially prioritizing profit over public interest. Therefore, considered and agile regulation is essential for encouraging responsible innovation while safeguarding end-users and vulnerable groups.

This section provides policymakers with a toolbox of AI governance instruments that they can use as a starting point for governing AI in their country contexts. For the purpose of this paper, we have identified 4 main types of regulatory approaches.

1. **Industry self-governance**
2. **Soft law (including technical standards)**
3. **Regulatory sandboxes**
4. **Hard law**

[36] https://www.un.org/sites/un2.un.org/files/un_ai_advisory_body_governing_ai_for_humanity_interim_report.pdf%20at%20, at p.8.

[37] Id.

[38] WDR 2021.

[39] See e.g. https://www.ohchr.org/en/press-releases/2019/10/world-stumbling-zombie-digital-welfare-dystopia-warns-un-human-rights-expert?LangID=E&NewsID=25156 %20https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4753107.

ABNASIA.ORG

For each of these regulatory tools, we have included examples from specific country contexts, discussing each tool's relative strengths, weaknesses, and policy tradeoffs. A summary of our analysis is set out in table 1.

This overview is not intended to be a comprehensive or exhaustive list of all AI governance interventions (given that such a list would quickly become outdated). The aim here is to provide an overview of current thinking around key tools to stimulate high-level policy debate.

This paper does not aim to set out a prescriptive list of 'best practices' – given the nascent state of AI governance and regulation efforts globally, it is too early to definitively state that some approaches work better than others. Instead, this paper instead aims to provide a toolbox of potential options that policymakers can consider and adapt for their local contexts.

It is also important to note that these regulatory tools are not discrete or standalone approaches – these approaches do not operate in isolation; they are interdependent and mutually reinforcing, and often intersect with other legacy regulatory and policy frameworks, both horizontal and sector-specific. Effective national AI governance strategies will likely integrate multiple tools.

Therefore, there is no 'one-size-fits-all' AI governance approach. Each tool has its own context-specific strengths and weaknesses. Policymakers should tailor any regulatory tool for their country's policy priorities and for the needs of local communities to create an AI governance regime that suits their national policy objectives. It is imperative that policymakers do not import regulatory provisions or strategies from other countries without appropriate modifications and consultation with affected communities, the public, civil society, the private sector, and the international community.

The tools outlined below are intended to apply to all AI systems – however, certain interventions (e.g. AI Safety Institutes) are particularly tailored to the governance of frontier, advanced large-scale AI systems. These interventions will be flagged for the reader as needed.

ABNASIA.ORG

Table 1. Governance Tradeoffs for AI Governance

Regulatory Tool	Examples	Benefits	Risks
Industry self-governance			
Private ethical codes and councils	• Microsoft Aether Committee and Responsible AI Standard Playbook • Google AI Principles • Bosch Ethical Guidelines for AI • IBM's AI Ethics Board • Partnership on AI (non-profit coalition on AI)	1. Can directly impact AI practices if integrated into business models and company cultures 2. Requires minimal public sector supervision, intervention or resources to set up	1. May be vague and of limited practical use. 2. Not appropriate for certain sectoral use-cases with particularly high risks – e.g. financial sector or healthcare. 3. Non-binding, with no mechanisms for effective public oversight or enforcement 4. Limited public input into design or implementation 5. Risk of 'ethics-washing,' where ethical commitments are superficial 6. Limited to a smaller subset of companies
Soft Law			
Non-binding international agreements	• OECD/G20 AI Principles • UNESCO Recommendation on the Ethics of AI • G7 Principles • UN General Assembly resolution on AI	1. Can directly impact national AI policy, when supported with funding and technical advice 2. Can have a global harmonizing effect	1. Non-binding 2. Focus on high-level principles rather than specific rights and responsibilities 3. Potential legal uncertainty due to vagueness/lack of practical impact
National AI principles / ethics frameworks	• UK AI regulation principles (2023 white paper) • US White House AI Bill of Rights • Australia voluntary AI Ethics Principles • Singapore Model AI Governance Framework for Generative AI	1. Provides guidance for industry actors 2. Agile and flexible; can adapt to technological advances 3. Relatively low-cost to create and promote	1. Non-binding 2. Potential legal uncertainty due to lack of clarity and practical implications. 3. Must be supported by mandatory transparency requirements to monitor uptake
Technical standards	• IEEE P70xx series • ISO/IEC 23894:2023 • NIST AI Risk Management Framework • UK AI Standards Hub • C2PA standards	1. Provides technical means of operationalizing responsible AI principles 2. Often have strong incentives for compliance 3. Usually created through multi-stakeholder process	1. Well-resourced incumbents could have disproportionate influence 2. Participation gaps for less developed states and civil society 3. Time-intensive to develop

ABNASIA.ORG

Regulatory Tool	Examples	Benefits	Risks
Regulatory sandboxes			
Regulatory sandboxes	• Colombia regulatory sandbox on privacy by design and default in AI projects • Brazil regulatory sandbox pilot for AI and data protection • Singapore AI Verify toolkit	1. Controlled environment to test and evaluate new regulatory approaches 2. Can leverage expertise of existing supervisory authorities 3. Collaborative form of regulation particularly suited for nascent AI ecosystems with limited capacity	1. Mainly useful where there are regulatory questions that can be solved by experimentation 2. Extremely resource-intensive 3. Can create market distortion and unfair competition
Hard law			
New horizontal AI law	• EU AI Act • Council of Europe Framework Convention • Brazil AI Bill • Chile AI Bill	1. Creates legal certainty and level playing field. 2. Sets binding, consistent level of protection against AI risks 3. Allows setting 'red lines' around unacceptable AI use cases	1. Lack of concrete 'best practices': policymakers should not 'copy and paste' approaches from other jurisdictions 2. Time-consuming and resource-intensive to design and implement 3. Tradeoffs in drafting (future-proofing vs. avoiding gaps in consumer protections)
Update or apply existing laws	• Data protection/ privacy • Human rights, equality, non-discrimination laws • Cybercrime • Intellectual property • Competition/antitrust • Procurement	1. Leverages existing regulatory architecture 2. Existing regulated entities already familiar with compliance framework	1. Limited by the scope of existing frameworks (e.g. data protection only applies to personal data). 2. Patchwork approach to regulation can create gaps in consumer protections and lack of legal certainty for industry
Targeted / sectoral laws or regulations	• Chinese regulations on recommendation algorithms, 'deep synthesis' technologies, and generative AI • New York City Local Law 144 of 2021 on Automated Employment Decision Tools • US semiconductor export controls	1. Can provide highly context-specific and sticky form of regulation 2. Particularly effective when enforced by existing sectoral regulators	1. Can create fragmented legal landscape, creating legal uncertainty and gaps in consumer protections 2. Risk becoming out of date if technological developments create new AI harms that do not map onto existing taxonomies

Source: Authors

ABNASIA.ORG

Tool 1: Industry Self-Governance

Private ethical codes and councils

There are a range of AI ethics documents and councils that have been set up by large technology firms or affiliated organizations.

Some are internal-facing, such as Microsoft's Aether Committee and Responsible AI Standard Playbook,[40] Google's AI Principles,[41] Bosch Ethical Guidelines for AI[42] and IBM's AI Ethics Board.[43] Other bodies, such as the Partnership on AI (established by Amazon, Apple, Google, Facebook, IBM, and Microsoft in 2016) aim to coordinate responsible AI work across industry, academia and civil society.

[40] https://www.microsoft.com/en-gb/ai/responsible-ai

[41] https://ai.google/responsibility/principles/

[42] https://www.bosch.com/stories/ethical-guidelines-for-artificial-intelligence/

[43] https://www.ibm.com/impact/ai-ethics

ABNASIA.ORG

Box 11: Partnership on AI (PAI)[44]

PAI is a multi-stakeholder nonprofit organization dedicated to the ethical and responsible development of artificial intelligence. Founded in 2016 and funded by philanthropic and corporate entities, PAI includes participation from technology companies, non-profits, and academic institutions.

Mission and Objectives

- *Responsible AI Development:* Ensuring AI technologies are ethical, transparent, and inclusive.
- *Interdisciplinary Collaboration:* Bringing together experts from computer science, ethics, law, and social sciences to address AI's challenges.
- *Public Awareness and Education:* Enhancing public understanding of AI, its impacts, and ethical considerations.
- *Best Practices and Guidelines:* Creating guidelines to promote fairness, accountability, and transparency in AI development.

Key contributions through its collaborative efforts include:

1. *Ethical Guidelines and Best Practices:* Developed and disseminated ethical guidelines and best practices for AI development and deployment.
2. *Research and Reports:* Published numerous studies on critical AI issues like bias, safety, privacy, and societal impacts, providing insights for policymakers and practitioners.
3. *AI Policy and Advocacy:* Been active in advocating for sound AI policies at both national and international levels.
4. *Working Groups:* PAI has several working groups focused on specific areas such as AI and labor, safety-critical AI, fair, transparent, and accountable AI, and social and societal influences of AI.
5. *Public Awareness and Education:* Raised awareness and educated the public on ethical AI through events, workshops, and initiatives.
6. *AI Incident Database:* Launched a database for collecting and analyzing AI incidents to improve safety and reliability.

While PAI has encouraged large multi-stakeholder dialogue on the responsible development and use of artificial intelligence, some have voiced concerns regarding the dominance of Big Tech in its activities, to the detriment of other actors: in 2020 prominent civil society organization Access Now resigned from PAI, citing a **lack of consensus and radically differing views between stakeholders' and 'an increasingly smaller role for civil society to play within PAI', stating that they 'did not find that PAI influenced or changed the attitude of member companies or encouraged them to respond to or consult with civil society on a systematic basis.**[45]

Source: https://partnershiponai.org/; https://www.accessnow.org/press-release/access-now-resignation-partnership-on-ai/

[44] https://partnershiponai.org/

[45] https://www.accessnow.org/press-release/access-now-resignation-partnership-on-ai/

ABNASIA.ORG

Ethical codes and councils can be important governance instruments if they are directly integrated into the business models and company cultures of industry actors – providing a focal point for live, product-relevant questions on AI ethics.[46] They also require minimal public sector supervision or interventions (although governments can encourage the creation of such councils through law or regulatory guidelines).[47]

However, policymakers should also note their weaknesses. First, even if properly integrated into key product decisions on AI development, some principles and ethics documents may be too vague and therefore of limited practical use.[48] Second, because of their nature as non-binding guidelines, there are no mechanisms for effective public oversight or enforcement, with little transparency or public input into how these ethical guidelines are created or implemented.[49] Third, these ethical frameworks are often inconsistently interpreted and implemented;[50] there is a risk that industry stakeholders will engage in regulatory arbitrage by 'shopping' around for the most permissive ethical principles to allow for minimal interruption to business.[51]

For these reasons, self-governance will rarely be a standalone intervention. Given the potential risk of 'ethics-washing', the existence of such ethical principles should not be seen as a complete regulatory intervention and should not preclude further action by policymakers. Even where self-regulation is an appropriate governance intervention, regulators may still have a role to play in providing incentives or guidelines for responsible action.

[46] https://cms.law/en/media/local/cms-cmno/images/other/artificial-intelligence-what-is-an-ai-ethics-board-cms?v=1

[47] E.g. The creation of ethical councils may satisfy the recommendation under Article 29 Data Protection Working Party Guidelines on Automated individual decision-making and profiling of 3 October 2017, which advises data controllers to 'establish ethical review boards to assess the potential harms and benefits to society of particular applications for profiling.'. https://cms.law/en/media/local/cms-cmno/images/other/artificial-intelligence-what-is-an-ai-ethics-board-cms?v=1.

[48] Munn (2022), https://link.springer.com/article/10.1007/s43681-022-00209-w.

[49] https://www.annualreviews.org/content/journals/10.1146/annurev-lawsocsci-020223-040749, p.258.

[50] id.

[51] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3835010.

ABNASIA.ORG

Tool 2: Soft Law

Non-binding international agreements

Some of the earliest non-binding, country-led instruments on AI governance were adopted in intergovernmental fora such as the OECD, G20 and UNESCO. A brief timeline summarizing the key agreements and developments in this area are set out below:

Source: G20 Ministerial Statement on Trade and Digital Economy, G20, 3–4 (2019) https://www.mofa.go.jp/files/000486596.pdf; UNESCO (2021) https://unesdoc.unesco.org/ark:/48223/pf0000386276; https://unesdoc.unesco.org/ark:/48223/pf0000385198; https://digital-strategy.ec.europa.eu/en/news-redirect/800511; https://www.reuters.com/technology/britain-publishes-bletchley-declaration-ai-safety-2023-11-01/; https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf?token=yC0PshhQyLqa21sZ7w&fe=true

1

May 2019: OECD AI Principles

The OECD AI Principles are adopted, focusind on inclusive growth, sustainable development and well-being, human-centered values and fairness, transparency and explainability, robustness, security and safety, and accountability. These principles have been endorsed by 42 countries, including all OECD member countries and several others, making them one of the most widely recognized AI frameworks globally.

2

September 2019:G20 AI Principles

G20 AI Principles are adopted during the 2019 summit in Osaka, drawing heavily from the OECD's work. These principles emphasize human-centered values, fairness, transparency, and accountability.

3

November 2021:UNESCO Recommendation

UNESCO's 193 Member States adopt the Recommendation on the Ethics of Artificial Intelligence, with accompanying readiness and impact assessment guidance

4

October 2023:G7 AI Principles and Code of Conduct

G7 countries announce a set of international guiding principles on artificial intelligence and a voluntary code of conduct for AI developers.

5

November 2023:AI Safety Summit and Bletchley Declaration

UK government convenes the world's first AI Safety Summit, leading to the signing of the Bletchley Declaration. This declaration aims to boost global efforts to cooperate on AI safety, focusing on identifying AI risks of shared concern, building scientific understanding on AI risks, and developing cross-country policies to mitigate such risks.

March 2024:UN Resolution on AI

UN General Assembly adopts a landmark resolution on the promotion of "safe, secure and trustworthy" artificial intelligence. The resolution emphasizes the need to respect, protect and promote human rights in the design, development, deployment, and the use of AI, while also recognizing AI systems' potential to accelerate and enable progress towards reaching the 17 Sustainable Development Goals.

7

May 2024:OECD AI Principle Updated

OECD AI Principles updated.

ABNASIA.ORG

Although these agreements are non-binding in nature, they demonstrate a notable degree of international consensus around key responsible AI principles. They can also have a direct impact on national AI policies – The OECD's 2023 report on the state of implementation of its AI principles has found that countries have sought to translate the AI Principles into concrete policy interventions through a range of measures, including 'i) establishing ethical frameworks and principles, ii) considering hard law approaches, iii) supporting international standardization efforts and international law efforts [...] and iv) promoting controlled environments for regulatory experimentation'.[52]

The UN's High-Level Advisory Body on AI (HLAB) has recognized the critical need for robust socio-technical standards to govern AI. In its interim report, the HLAB emphasized the importance of a coordinated global approach to prevent fragmentation and ensure interoperability among various AI governance frameworks.[53] It calls for inclusive participation, especially from the Global South, and underscores the importance of aligning AI governance with international human rights laws. The report also highlights AI's potential to achieve the Sustainable Development Goals (SDGs) through ethical and inclusive deployment. The final report is expected before the end of 2024.

However, it is important to recognize that these agreements are not standalone regulatory interventions – they focus mainly on high-level principles and do not address important questions regarding the assignment of rights and regulatory responsibilities. For these documents to have practical relevance, they must be translated into national policy frameworks and accompanied by further technical assistance and policy advice. Often, international organizations will provide direct technical assistance to member states to help guide their AI policy development – for example, it was announced in May 2024 that Chile had adopted an updated national AI policy and action plan, following the recommendations of a Readiness Assessment Report elaborated by UNESCO (see Box 12).[54]

[52] https://www.oecd-ilibrary.org/docserver/835641c9-en.pdf?expires=1716551804&id=id&accname=ocid195787&checksum=8B861C98822A13F96A39FDF353858786, p. 15.

[53] https://www.un.org/sites/un2.un.org/files/un_ai_advisory_body_governing_ai_for_humanity_interim_report.pdf.

[54] https://www.unesco.org/en/articles/chile-launches-national-ai-policy-and-introduces-ai-bill-following-unescos-recommendations#:~:text=The%20country%2C%20following%20the%20recommendations,responsible%20development%20of%20this%20technology

ABNASIA.ORG

Box 12: Chile-UNESCO Collaboration on AI Policy – UNESCO Readiness Assessment Methodology[53]

Chile was one of the first countries in the world to implement and finalize UNESCO's Readiness Assessment Methodology (RAM). The RAM is intended to help countries understand how prepared they are to implement AI ethically and responsibly for their consumers, while also highlighting what further institutional and regulatory changes are needed.[53]

The implementation of a RAM has three stages:

1. Diagnosis of national AI landscape
2. Development of national AI multi-stakeholder roadmap
3. Main policy recommendations for national AI strategy

During June and July 2023, participatory consultations were held with different actors in the local AI ecosystem, with the aim of generating recommendations for AI development in Chile. The Chilean Ministry of Science, Technology, Knowledge and Innovation (MSTKI), in collaboration with UNESCO, identified six thematic areas of discussion relevant to the AI agenda for the coming years, covering the future of work, democracy, government, health, education, safety, regulation and the environment.

Chile's engagement with UNESCO was led by MSTKI, the ministry that elaborated Chile's 2021 National AI Policy. MSTKI was supported by a Ministerial Steering Committee that included MSTKI, the Ministry of Economy, Development and Tourism, and the Ministry of Education.

In each area of discussion, participants were asked to identify challenges and opportunities; the outcomes of these discussions served as inputs for the main recommendations of the Readiness Assessment Report. The final Report recommended the following list of ten recommendations, to align Chile's AI policy with UNESCO's recommendations:[54]

1. REGULATION
    1.1. Assign urgency to the updating of the current Personal Data Protection Law and the Cybersecurity and Information Critical Infrastructure Bill
    1.2. Create a multi-stakeholder and adaptive governance for AI regulation
    1.3. Explore Regulatory Experimentation Mechanisms (e.g., Sandboxes) for the Application of AI in Critical Areas
    1.4. Promote ethical principles of AI through purchasing regulations and standards
2. INSTITUTIONAL FRAMEWORK
    2.1. Improve data collection and statistics on the use of AI
    2.2. Development of AI Strategies for Local Governments
    2.3. Update Chile's National AI Policy (NAIP)

(Continued on other page)

https://www.unesco.org/en/articles/chile-launches-national-ai-policy-and-introduces-ai-bill-following-unescos-recommendations#:~:text=The%20AI%20bill%20introduced%20by,the%20protection%20of%20consumers%20from; https://dig.watch/updates/chile-introduces-updated-national-ai-policy-and-new-ai-legislation#:~:text=Chile%20has%20officially%20launched%20its,Readiness%20Assessment%20Report%20by%20UNESCO

https://unesdoc.unesco.org/ark:/48223/pf0000387216

ABNASIA.ORG

(Box 12 continued)

3. CAPACITY BUILDING

    3.1. Development of Human Capital in AI

    3.2. Attract investments in AI technological infrastructure and promote discussion on its environmental impacts.

    3.3. Assess the impact of AI and automation on the workforce and define job retraining plans

To implement the findings from the RAM, in May 2024 Chile launched its updated National AI Policy and action plan, along with a proposed AI bill spearheaded by MSTKI, seeking to regulate and encourage the ethical and responsible development of AI.[57] The Bill sets out a risk-based approach to regulation (for more information on a risk-based approach to AI regulation, see section [x] below) classifying AI systems into unacceptable, high, limited and no evident risk categories.[58] Chile's revised National AI Policy explicitly incorporated insights from the RAM process, addressing governance gaps and integrating diverse stakeholder perspectives from across Chile.

Source: https://www.unesco.org/en/articles/chile-launches-national-ai-policy-and-introduces-ai-bill-following-unescos-recommendations#:~:text=The%20AI%20bill%20introduced%20by,the%20protection%20of%20consumers%20from; https://dig.watch/updates/chile-introduces-updated-national-ai-policy-and-new-ai-legislation#:~:text=Chile%20has%20officially%20launched%20its,Readiness%20Assessment%20Report%20by%20UNESCO

National AI principles / ethics frameworks

In addition to ethical frameworks and principles-based documents created by industry actors, and international bodies, governments are increasingly developing voluntary national AI principles and ethics frameworks (as noted in Box 12 above on Chile's national AI policy, adopted with assistance from UNESCO, there is often a significant interplay between international and national frameworks).

For example, a 2023 white paper released by the UK sets out 5 principles : 1) Safety, security and robustness, 2) appropriate transparency and explainability, 3) fairness, 4) accountability and governance, 5) contestability and redress) to guide the responsible development and use of AI across all sectors.[60] In 2022 the Office of Science and Technology Policy (OSTP) of the White House produced a 'Blueprint for an AI Bill of Rights' suggesting fundamental principles to guide and govern the efficient development and implementation of AI systems while in Rwanda, Guidelines on the Ethical Development and Implementation of Artificial Intelligence[61] was released as part of the National AI Strategy.

The key characteristic of these frameworks is that they are non-binding. For instance, the UK white paper explains that these principles were not placed on statutory footing to allow the government to remain agile and respond quickly and proportionately to new

https://unesdoc.unesco.org/ark:/48223/pf0000387216

https://www.unesco.org/en/articles/chile-launches-national-ai-policy-and-introduces-ai-bill-following-unescos-recommendations#:~:text=The%20AI%20bill%20introduced%20by,the%20protection%20of%20consumers%20from

[59] id.

[60] https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#:~:text=Our%20framework%20is%20underpinned%20by,Fairness

[61] https://www.minict.gov.rw/index.php?eID=dumpFile&t=f&f=67550&token=6195a53203e197efa47592f40ff4aaf24579640e

ABNASIA.ORG

technological advances.[62] Similarly, Australia has adopted voluntary AI Ethics Principles[63] to guide the development and deployment of AI technologies. These principles provide a flexible framework to address ethical considerations without imposing mandatory regulations, allowing for rapid adaptation as AI technology evolves. Both examples illustrate how non-binding frameworks can serve as interim measures, providing industry guidance while preserving the ability for future regulatory adjustments based on emerging insights and risks.

National AI principles and ethical frameworks can form useful intermediate stopgaps as part of a broader 'wait and see' approach to AI regulation but they must be carefully monitored. Non-binding frameworks provide a useful guide for the industry thereby promoting responsible innovation. However, it is important to note that any 'wait and see' approach must be carefully implemented, and any period of 'active learning' may need to be supported by strong transparency requirements to allow policymakers to actively monitor AI developments and ensure that consumers and vulnerable groups are not exposed to unacceptable levels of risk. A 'wait and see' approach should not preclude further action, whether that is in the form of hard regulation, development of national AI standards, or implementation of a regulatory sandbox. For example, the UK government has said that it eventually expects to introduce 'targeted, binding requirements' for the most powerful general-purpose AI systems.[64]

Technical standards and certification frameworks

Voluntary international standard-setting organizations are increasingly developing technical standards[65] for AI governance. One early example is the Institute of Electrical and Electronics Engineers' (IEEE) P70xx series of standards for ethical use of AI. Known as the 'Ethically Aligned Design' (EAD) principles, they include standards on transparency (7001–2021), processes for considering ethical issues in design (7000–2021), and standards on bias and 'ethically-driven nudging' (7003TM, 7008TM). The International Organization for Standardization (ISO) is also increasingly active in this area and has published a standard on AI risk management (ISO/IEC 23894:2023).

Similar work is being undertaken by national standards institutes such as the U.S. National Institute of Standards and Technology (NIST), whose AI Risk Management Framework (RMF) provides comprehensive guidance for risk mitigation across the AI lifecycle.[66] The UK has also developed an AI Standards Hub to share knowledge, capacity, and research on AI standards.[67] There are also standard-setting bodies that seek to address specific governance issues such as misinformation: for example, the Coalition for Content Provenance and Authenticity (C2PA), which includes stakeholders such as Adobe, BBC, Google, Microsoft, Sony and OpenAI, seeks to address online misleading information through developing technical standards that certify the source and history (or provenance) of media content.[68]

[62] https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#:~:text=Our%20framework%20is%20underpinned%20by,Fairness

[63] https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles

[64] https://www.gov.uk/government/news/uk-signals-step-change-for-regulators-to-strengthen-ai-leadership

[65] This paper defines 'technical standards' as technical specifications that encourage (but do not require) compliance. Historically, technical standards have been crucial for the development of the internet and other networked infrastructures – one important factor driving standard adoption is the need for interoperability.

[66] https://www.nist.gov/itl/ai-risk-management-framework

[67] https://aistandardshub.org/

[68] https://c2pa.org

ABNASIA.ORG

Figure 2. AI Standards Landscape Snapshot
Source: AI Standards Hub, Q2 2024

Technical standards can be complemented by certification schemes, trust marks, quality marks and seals. Some examples include the proposed AI certification 'Made in Germany,'[69] IEEE's Ethics Certification Program for Autonomous and Intelligent Systems,[70] the proposed Malta's National AI Certification Framework,[71] Responsible Artificial Intelligence Institute Certification,[72] and Denmark's digital trust seal.[73]

Although many of these standards address 'ethical' issues, they are distinct from private ethical principles in two ways. First, high-level ethical principles will often recognize the importance of key principles such as transparency in AI systems, without specifically elaborating how they can be implemented in context. In contrast, technical standards operate at a greater level of detail, seeking to explain how such principles can be integrated into AI systems in practice – the IEEE Standard for Transparency of Autonomous Systems (IEEE 7001-2021), for example, directly sets out a methodology for creating **measurable, testable levels of transparency, so that autonomous systems can be objectively assessed, and levels of compliance determined.**[74] Second, although technical standards are non-binding in nature, there are often strong incentives for compliance – especially where adherence to a standard becomes necessary for interoperability, or if a certification becomes a de facto industry benchmark. AI standards may play a crucial role in 'regulatory interoperability' across borders – as different countries enact AI legislation. Differences in regulatory language and approaches to key principles such as trustworthiness, accountability, and transparency can create obstacles for regulatory compliance by industry actors.

Technical standards also have a role to play in creating multi-stakeholder driven global consensus on how these principles are translated into technical specifications.[75] Standard development is often a multi-stakeholder process and can incorporate input from the technical community, governments, academia, and civil society organizations. In short, standards can provide an agile way of translating responsible AI

[69] https://www.ki.nrw/en/flagships-en/certified-ai/
[70] https://standards.ieee.org/industry-connections/ecpais/
[71] https://www.mdia.gov.mt/malta-ai-strategy/
[72] https://www.responsible.ai/
[73] https://d-seal.eu/
[74] https://standards.ieee.org/ieee/7001/6929/
[75] https://www.holisticai.com/blog/ai-governance-risk-compliance-standard

ABNASIA.ORG

practices to the technical level while also facilitating a multi-stakeholder approach to global harmonization – although this greatly depends on the participation structure of the relevant standard-setting organizations.

Table 2.

Standard	Coverage	Responsible Institute
ISO/IEC 22989:2022	Framework for AI, addressing AI concepts, terminology, and principles.	International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC)
ISO/IEC 23053:2022	Framework for AI systems, focusing on machine learning lifecycle processes.	ISO and IEC
IEEE 7000-2021	Model process for addressing ethical concerns during system design.	Institute of Electrical and Electronics Engineers (IEEE)
IEEE 7010-2020	Well-being metrics for ethical AI and autonomous systems.	IEEE
NIST AI Risk Management Framework (AI RMF)	Guidelines for organizations to identify and managing risks associated with AI technologies, focusing on accuracy, reliability, and robustness.	National Institute of Standards and Technology (NIST)
BS 8611:2016	Guide to the ethical design and application of robots and robotic systems.	British Standards Institution (BSI)
ISO/IEC JTC 1/SC 42	Comprehensive AI standard covering terminology, data quality, risk management, and governance.	ISO/IEC Joint Technical Committee 1/ Subcommittee 42
IEEE P70xx Series	Standards for ethical use of AI including transparency (7001-2021), ethical design (7000-2021), and bias (7003TM, 7008TM).	IEEE
ISO/IEC 23894:2023	AI risk management standard focusing on managing risks throughout the AI lifecycle.	ISO and IEC
C2PA Technical Standards	Standards addressing online misinformation by certifying the source and history of media content.	Coalition for Content Provenance and Authenticity (C2PA)
Responsible AI Institute Certification	Certification framework for assessing AI systems against ethical and technical standards.	Responsible AI Institute
Malta's National AI Certi-fication Framework	National framework for certifying AI systems in accordance with ethical and technical standards.	Government of Malta
IEEE Ethics Certification Program for Autonomous and Intelligent Systems	Certification for AI systems based on ethical standards and transparency.	IEEE
Denmark's Digital Trust Seal	Certification mark for trustworthy AI systems focusing on transparency and accountability.	Danish Government

ABNASIA.ORG

However, although many standard-setting organizations adopt a multi-stakeholder approach to standards development, the most well-resourced industry players with technical expertise and financial resources often have an advantage in these processes. These large actors can sometimes coerce other actors into adopting certain standards.[76] This can also lead to participation gaps for less well-resourced actors, such as less developed countries and civil society organizations (see Box 13 below for more detail).[77] In addition, **the multi-stakeholder and consensus-based process of some standard-setting organizations means that standards-development can take time** – for example, developing an ISO standard from first proposal to final publication usually takes around 3 years.[78] This can pose a challenge for agile governance given how quickly AI systems are evolving.

Box 13: Participation Gaps at Standard-Setting Organizations

Governments can directly participate in standard-setting processes where possible. However, participants in these processes are often technical representatives usually from industry, typically sponsored by a handful of large private sector actors based in the Global North. Among the important standard-setting bodies governing the ICT sector, only the International Telecommunication Union (ITU) has express provisions for participation by countries from the Global South.[79]

Civil society organizations, such as the Ada Lovelace Institute, have identified significant barriers to participation in AI standardization processes. These barriers include the substantial time commitment required, the complexity and opacity of the processes, and the dominance of industry voices, which can marginalize less-resourced actors and civil society groups.[80] For instance, the EU standardization body tasked with creating standards for 'high-risk' AI systems under the EU AI Act faces challenges in ensuring broad stakeholder participation.

Given these participation issues, governments have two primary ways to engage with AI standards regimes:[81]

1. **Hybrid Approach:** This method involves specifying that compliance with certain standards satisfies legal obligations. For instance, the EU AI Act adopts this approach by requiring providers of 'high-risk' AI systems to self-certify that they meet essential requirements set out in proprietary standards authored by the European Committee for Standardization (CEN) and the European Committee for Electrotechnical Standardization (CENELEC). This gives these standardization bodies significant regulatory influence globally, as AI providers wishing to sell high-risk systems in the European market must certify their compliance with these standards.[82] *(for more information on the EU AI Act's risk based regulatory approach, see BOX 14)*
2. **Symbiotic Approach:** In this approach, a legal regime promotes optional industry certification mechanisms. An example is the EU data protection law, which encourages companies to adhere to certain standards voluntarily, thus fostering a culture of compliance through incentives rather than mandates.

Source: Kanevskaia (2023), p.263-267, https://www.cambridge.org/core/books/law-and-practice-of-global-ict-standardisation/069342911A73905590EEA61655CA0DA0, https://www.adalovelaceinstitute.org/report/inclusive-ai-governance/; Veale (2023), p.262.

[76] https://www.annualreviews.org/content/journals/10.1146/annurev-lawsocsci-020223-040749, p.262.

[77] https://ideas.repec.org/a/eee/telpol/v45y2021i6s0308596121000483.html

[78] https://www.iso.org/developing-standards.html#:~:text=The%20voting%20process%20is%20the,usually%20takes%20about%203%20years.

[79] p.263-267, https://www.cambridge.org/core/books/law-and-practice-of-global-ict-standardisation/069342911A73905590EE61655CA0DA0

[80] https://www.adalovelaceinstitute.org/report/inclusive-ai-governance/

[81] https://www.annualreviews.org/content/journals/10.1146/annurev-lawsocsci-020223-040749, p.262.

[82] https://www.annualreviews.org/content/journals/10.1146/annurev-lawsocsci-020223-040749, p.264.

ABNASIA.ORG

Tool 3: Hard Law

A growing number of countries have enacted binding legislation establishing concrete obligations and consequences for AI development and use. These hard laws can take various forms, including horizontal laws that apply broadly across all sectors, technology-specific laws targeting particular types of AI applications or systems, or sector-specific laws addressing AI deployment within specific industries.

Given the speed at which new AI laws are being proposed, this paper does not provide a comprehensive survey of all proposed approaches (given that this would quickly become out of date) – instead, we group current regulatory proposals into several loose categories, to identify common strengths, weaknesses, and policy tradeoffs.

According to Stanford University's AI Index report, 31 countries have passed at least one AI-related bill since 2016.[83]

Creation of horizontal AI law

This section focuses on 'horizontal' AI laws, that apply to all AI systems[84] regardless of sector or use case.

A 'risk-based' approach involves categorizing AI applications based on their potential risks and impacts. This approach subjects higher-risk AI systems to more rigorous regulatory obligations to mitigate potential harms.[85] The EU AI Act, passed in 2024, is one of the world's first examples of such a law – its approach draws heavily from EU product liability law, classifying AI systems according to their risk levels and applying tailored regulatory requirements accordingly. It also applies more stringent requirements to developers of 'general-purpose AI' models and imposes additional requirements for those posing 'systemic' risks. A deeper evaluation of the EU's risk-based approach is set out in box 14.

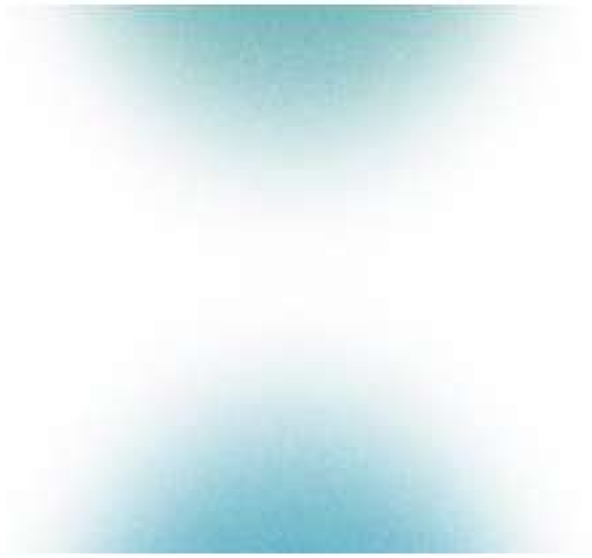


Figure 3.

Source: https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report-2023_CHAPTER_6-1.pdf

[83] https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report-2023_CHAPTER_6-1.pdf

[84] It is important to note that jurisdictions have taken different approaches to defining 'AI' and 'AI systems' in their proposed legislative frameworks. For example, law firm White and Case notes that 'the draft text of the EU AI Act adopts a definition of 'AI systems' that is based on (but is not identical to) the OECD's definition, and which leaves room for substantial doubt due to its uncertain wording'. https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker#home

[85] Digital Trends Report 2023, WBG

Box 14: Evaluating the EU AI Act's risk-based regulatory framework

First introduced in 2021, the EU AI Act is a legislative proposal by the European Commission that introduces a tiered, risk-based approach to regulating AI within the European Union.

The Act classifies AI systems into four risk categories: unacceptable risk, high risk, limited risk, and minimal risk.

- **Unacceptable risk:** AI systems that pose a clear threat to safety, livelihoods, or rights, such as social scoring by governments, are banned. These can include things such as social scoring, manipulative AI, exploitation of vulnerable populations, biometric categorization or automated compiling of facial recognition databases.
- **High risk:** AI systems used in critical infrastructure, education, employment, law enforcement, and health must meet strict requirements before they can be placed on the market. The obligations and conformity assessments include areas such as risk and quality management systems, data governance, technical documentation, record-keeping, instructions for downstream deployers, and design for accuracy, robustness and cybersecurity.
- **Limited risk:** AI systems with 'limited risk' are subject to light transparency obligations such as ensuring that end-users are aware they are interacting with AI (e.g. chatbots and deepfakes must declare use of AI).
- **Minimal risk:** Most AI systems, like spam filters or video games, fall under this category and are largely unregulated.

In addition, all general-purpose AI model developers must provide technical documentation, instructions for use, comply with the EU Copyright Directive, and publish a summary about training data. Meanwhile, general purpose models posing 'systemic risks[86] face additional requirements – all providers of these models – whether open source or not – must also conduct evaluations of models for risks, adversarial testing, and track and report serious incidents and ensure cybersecurity protections.[87]

Because the EU AI Act is one of the first pieces of binding legislation regulating AI, policymakers designing new regulatory frameworks in other countries may seek to draw inspiration from it.[88] However, this approach should be treated with caution – the EU AI Act is grounded in concerns specific to the EU, such as the need to create a harmonized regulatory regime across 27 member states. The EU AI Act's risk categorization framework is therefore the product of the specific political compromise, as well as drafting specific to EU product liability and consumer protection law. Because this formulation may not reflect the policy priorities of other countries, the AI Act should not be 'copied-and-pasted' into new regulatory regimes without appropriate modifications

(Continued on other page)

[86] AI models pose systems risks where the cumulative amount of compute used for its training is greater than 10^{25} floating point operations (FLOPs) - similar to the level of OpenAI's GPT4 which powers ChatGPT)

[87] Id.

[88] A similar process occurred in the data protection context, leading to the 'Brussels effect' of the GDPR, https://academic-oup-com.libproxy-wb.imf.org/book/36491?login=true&token=

ABNASIA.ORG

(Box 14 continued)

Policymakers should also note the following issues that have been highlighted by academics and civil society regarding the EU AI Act:

1. AI systems covered under the AI Act are those that 'may exhibit adaptiveness after deployment'.[89] This definition could potentially exclude AI systems that do not learn or adapt to new data inputs after deployment, such as older rule-based systems. However, these AI systems may still be complex and cause unique risks for consumers.
2. The AI Act is the product of a unique blend of EU product safety regulation, fundamental rights protection, and consumer protection law. Academics have argued that this patchwork approach to regulation leaves certain gaps in how the Act is drafted.[90] In addition, because the Act acts as a form of 'maximum' market harmonization under EU law, in principle member states cannot introduce further national regulation on AI.[91]
3. Many of the 'essential requirements' that high-risk AI systems must comply with are drafted in general and vague terms (e.g. they must have an 'appropriate level of accuracy, robustness, and cybersecurity' to mitigate risks to fundamental rights[92]). The AI Act therefore relies on European standards development bodies (particularly CEN-CENELEC) to clarify these essential requirements by operationalizing them into harmonized European technical standards.[93] Under the AI Act, high-risk AI systems and general-purpose AI systems that are in conformity with these harmonized standards are automatically presumed to comply with the Act's legal requirements for high-risk systems.[94] However, as noted at box 13 above, these European standards development organizations face serious participation gaps and do not have specific expertise in important fundamental rights topics. More generally, academics have noted that the Act's approach to its enforcement architecture means that key operative provisions delegate critical regulatory tasks to AI providers themselves, without adequate oversight or redress mechanisms.[95]
4. The 'list-based' approach of the Act means that the Act may not guard against novel AI systems that fall outside the Act's risk classification.[96]
5. Persons affected by AI systems have no specifically enforceable rights or role within the AI Act[97] (although individual rights regarding automated decision-making are found elsewhere in EU law, such as under the GDPR).[98]

[89] Art. 3(1), EU AI Act.

[90] https://papers.ssrn.com/sol3/Delivery.cfm?delivery_id=3896852&frd=yes&anym=yes

[91] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4874852

[92] Art. 15, EU AI Act.

[93] https://www.adalovelaceinstitute.org/wp-content/uploads/2023/03/Ada-Lovelace-Institute-Inclusive-AI-governance-Discussion-paper-March-2023.pdf. This is an approach that is modelled on the EU's 'New Legislative Framework' market surveillance regime.

[94] Art. 40, EU AI Act.

[95] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4874852

[96] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4874852

[97] https://www.adalovelaceinstitute.org/report/regulating-ai-in-europe/

[98] Article 22(1) of the GDPR gives data subjects the 'right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.'

ABNASIA.ORG

Another approach to horizontal AI regulation is a 'rights-based' approach – one example is the Council of Europe's recently finalized Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law.[99] Countries that sign up to the Framework Convention commit to adopt or maintain measures to ensure that AI activities are compatible with human rights,[100] and to ensure that AI systems are not used to undermine the integrity of democratic processes and the rule of law.[101] Countries also commit to ensuring that their national frameworks (a) incorporate general principles regarding AI governance (transparency, accountability, equality, privacy, etc.),[102] (b) contain measures ensuring accessible and effective remedies for rights violations,[103] and (c) have mechanisms to assess and mitigate adverse AI impacts on rights.[104] However, the Convention has been criticized for the fact that countries are able to determine whether to apply the Convention to private sector actors, or implement 'other appropriate measures'.[105]

Brazil's proposed AI Bill[106] takes a hybrid approach – it is explicitly rights-based, but also incorporates a tiered risk-based model inspired by the EU AI Act; see box 15.

[99] The Committee on Artificial Intelligence (CAI), the body tasked with drafting the treaty, comprises the 46 member states of the Council of Europe, as well as observer states from most regions of the world: including Argentina, Canada, Costa Rica, the Holy See, Israel, Japan, Mexico, Peru the USA, and Uruguay, https://rm.coe.int/terms-of-reference-of-the-committee-on-artificial-intelligence-cai-/1680ade00f. Although developed under the auspices of the Council of Europe, the Convention is open to ratification by any country. The Council of Europe can have global influence: for example, the Budapest Convention on Cybercrime has 67 ratifications or accessions and is widely considered to be a key international instrument governing cybercrime.

[100] Article 4, Framework Convention.

[101] Article 5, Framework Convention.

[102] Chapter III, Framework Convention.

[103] Chapter IV, Framework Convention.

[104] Chapter V, Framework Convention.

[105] See wording at Art 3(1)(b), Framework Convention; for discussion of the legislative history of this private sector carveout, see https://www.euractiv.com/section/artificial-intelligence/news/eu-commissions-last-minute-attempt-to-keep-private-companies-in-worlds-first-ai-treaty/.

[106] Bill No. 2,338/2023 https://legis.senado.leg.br/sdleg-getter/documento?dm=9347593&ts=1683152235237&disposition=inline&_gl=1*adqokm*_ga*MTgyMDY0MTcwMS4xNjc5OTM2MTI0*_ga_CW3ZH25XMK*MTY4MzIxNzUzMy4yLjEuMTY4MzIyMDAyMy4wLjAuMA

ABNASIA.ORG

Box 15: Brazil's Bill 2.338/2023

Brazil's Bill 2.338/2023, currently under consideration, proposes a comprehensive risk and rights-based approach to AI governance. It defines an AI System as a '[c]omputer system, *with different degrees of autonomy, designed to infer how to achieve a given set of goals, … predictions, recommendations, or decisions that can influence the virtual or real environment.*'

Risk-Based Approach

The risk-based approach mandates that AI systems conduct a preliminary self-assessment analysis to classify themselves according to their risk levels. These levels include:

- *Prohibited AI Systems:* These are deemed excessively risky and are banned.
- *High-Risk AI Systems:* These systems can be used only if they meet stringent compliance requirements such as impact assessments, robustness, accuracy, reliability, and human oversight. High-risk AI systems cover applications like credit rating, personal identification, autonomous vehicles, medical diagnoses, and decision-making processes affecting employment, education, and access to essential services. Developers and operators of these systems must ensure they do not use AI for subliminal manipulation or exploit vulnerabilities of specific groups, such as children or people with disabilities. In addition, high-risk AI systems must include technical documentation, log registers, reliability tests, technical explainability measures and measures to mitigate discriminatory biases.[107]

Every AI system must implement a governance structure involving transparency, data governance and security measures.[108]

The Bill places significant emphasis on organizations' responsibility to mitigate biases through regular public impact assessments. These impact assessments will be held in an open public database.

Rights-Based Approach[109]

The Bill also proposes individual rights, such as the right to explanation about decisions, non-discrimination and correction of discriminatory biases, and the right to privacy and protection of personal data.[110] In addition, rules for civil liability, codes of best practice, notification of AI incidents, copyright exceptions for data mining processing, and fostering of regulatory sandboxes are also included.

To implement this, the bill proposes an institutional model with four coordinated bodies:

1. **The Competent Authority:** Likely the National Data Protection Authority (ANPD), responsible for interpreting and regulating AI law.
2. **The Executive Branch:** Formulates public policies for AI development and is tasked with designating supervisory authority to regulate and enforce legislation regarding Brazil's National AI Strategy (EBIA).
3. **Sectoral Regulatory Bodies:** Specific regulators working in cooperation with the ANPD.
4. **The Artificial Intelligence Advisory Council:** Ensures societal participation in AI-related decisions

Source: Bill 2.338/2023; https://oecd.ai/en/work/brazils-path-to-responsible-ai; https://accesspartnership.com/access-alert-brazils-new-ai-bill-a-comprehensive-framework-for-ethical-and-responsible-use-of-ai-systems/

[107] https://accesspartnership.com/access-alert-brazils-new-ai-bill-a-comprehensive-framework-for-ethical-and-responsible-use-of-ai-systems/ https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-brazil

[108] Id.

[109] https://oecd.ai/en/work/brazils-path-to-responsible-ai

[110] https://accesspartnership.com/access-alert-brazils-new-ai-bill-a-comprehensive-framework-for-ethical-and-responsible-use-of-ai-systems

ABNASIA.ORG

Horizontal laws allow policymakers to mitigate the legal uncertainty created by non-binding frameworks. A horizontal AI law, backed by strong supervisory and enforcement capacity, can play a critical role in creating trust in the AI economy, which then enables greater participation by consumers in digital life and more robust responsible innovation ecosystems. Legal frameworks which clearly specify what types of AI-enabled activities, and AI systems, are unacceptable allow firms greater certainty in ensuring their activities are fully regulatory compliant. The creation of regulatory frameworks with clear 'red lines' around unacceptable AI use cases is an approach that has been recommended by the human rights community in particular.[111] The need for binding horizontal regulation has been recognized by the international community.[112]

However, implementing horizontal laws comes with several challenges.

There is a significant risk of regulatory fragmentation, where different jurisdictions develop incompatible or conflicting regulations. **This can lead to overlapping requirements that create compliance burdens for businesses operating in multiple regions.** Such fragmentation and overlap can hinder the growth and scalability of local AI innovation ecosystems by increasing the complexity and cost of compliance.

Regulatory fragmentation and vague legal frameworks can also result in uneven protection against AI risks for consumers. Inconsistent regulations across different regions may mean that some consumers enjoy robust protections against AI risks, while others are left vulnerable due to weaker or less comprehensive regulatory frameworks. This disparity can undermine public trust in AI technologies and exacerbate social inequalities. For example, the Council of Europe Framework Convention has been criticized by Amnesty International for its high-level approach, which is seen as excessively vague. Critics argue that it does not provide sufficient detail regarding the concrete rights affected by AI, the specific AI-based practices that are incompatible with human rights, and the processes for conducting effective and binding human rights due diligence for AI developers and deployers.[113]

Furthermore, crafting binding regulation is inherently challenging due to the absence of established 'best practices' and limited supervisory or enforcement experience. Policymakers often look to existing frameworks like the EU AI Act for guidance, but this approach must be adapted to fit local contexts. Simply copying the EU AI Act's categorizations of AI systems without modifications may result in regulations that are ill-suited to the specific needs and circumstances of different jurisdictions.[114]

Creating binding legislation is also a time-consuming and resource-intensive process. Legislative drafters may seek to ensure that AI regulations are flexible enough to adapt to rapid technological advancements by using high-level wording or allowing for future interpretation by courts and regulators. While this can help prolong the relevance of the regulations, it also introduces significant legal uncertainty. Industry actors may be left uncertain about how their regulatory obligations will be interpreted and enforced,[115] which can disadvantage startups with fewer compliance resources and slow down the deployment of beneficial AI technologies.

Fixed red lines or categorizations, such as those set out in the Brazilian and EU frameworks, may become outdated as

[111] https://www.amnesty.eu/wp-content/uploads/2024/04/EUs-AI-Act-fails-to-set-gold-standard-for-human-rights.pdf

[112] See UN High-Level Advisory Board Interim Report

[113] https://www.amnesty.eu/wp-content/uploads/2024/04/Amnesty-International-Recs-draft-CoECAI-11042024.pdf

[114] A similar process occurred in the data protection context, leading to the 'Brussels effect' of the GDPR. https://academic-oup-com.libproxy-wb.imf.org/book/36491?login=true&token=

[115] https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker#articles

ABNASIA.ORG

technology evolves.[116] This necessitates the development of innovative and iterative regulatory approaches that allow legal frameworks to be updated efficiently without incurring significant time and resource costs. Policymakers must strike a balance between providing clear, enforceable rules and maintaining the flexibility needed to adapt to future technological changes.

At the same time, policymakers should note that certain well-accepted AI governance principles, such as transparency and accountability, can be readily specified in statute in ways that remain technology-neutral but provide necessary documentation, auditability, and answerability requirements. For example, the Canadian Directive on Automated Decision-Making 2019[117] imposes several requirements regarding accountability and transparency before Canadian federal institutions are permitted to deploy automated decision systems: these include mandatory algorithmic impact assessment, mandatory user-facing notice requirements, and obligations to provide meaningful explanations to affected individuals after a decision is made, release custom source code owned by the Government of Canada, and document the decisions of automated decision systems.

Update or application of existing laws

Another approach is to focus on updating or amending existing regulatory frameworks that may apply to activities in the AI ecosystem. The non-exhaustive list of existing legal frameworks that can be applied to the AI ecosystem is on the next page.

[116] Indeed, the EU AI Act required significant last-minute amendments to account for the emergence of generative AI. https://www.reuters.com/technology/behind-eu-lawmakers-challenge-rein-chatgpt-generative-ai-2023-04-28/

[117] https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592

ABNASIA.ORG

Table 3.

Existing Legal Framework	Examples of Application to AI Ecosystem
Data protection / privacy	In March 2024, Singapore's Personal Data Protection Commission (PDPC) issued Advisory Guidelines on the Use of Personal Data in AI Recommendation and Decision Systems,[118] providing organizations with clarity on the use of personal data at three stages of AI system implementation: (a) development, testing and monitoring, (b) deployment, and (c) procurement.[119] On March 30, 2023, the Italian Data Protection Authority (DPA) ordered OpenAI to stop the use of ChatGPT to process the personal data of Italian data subjects, on the grounds that there was a material risk that ChatGPT would breach the GDPR, mainly due to the use of personal data as a training set for ChatGPT in the absence of an adequate legal basis and without provision of a privacy notice.[120]
Human rights, equality, and non-discrimination laws	In April 2024 the UK's Equality and Human Rights Commission (EHRC) issued a reminder to employers to prevent inadvertent bias or discrimination in their use of AI tools, following a complaint from an Uber Eats driver who argued that AI facial recognition checks required to access the Uber Eats platform were racially discriminatory.[121] In the US, the National Fair Housing Alliance (NFHA) and the US Department of Housing and Urban Development separately sued Facebook on the grounds that Facebook was allowing advertisers seeking to place algorithmic housing ads to exclude certain users by their race, which appeared to violate the US Fair Housing Act. The case was settled out of court by Meta.[122]
Cybercrime	In Nigeria, legal commentators have proposed the use of the Cybercrime (Prohibition, Prevention etc.) Act 2015 to combat deepfakes, via its prohibition on identity theft and impersonation.[123]

[118] https://www.pdpc.gov.sg/guidelines-and-consultation/2024/02/advisory-guidelines-on-use-of-personal-data-in-ai-recommendation-and-decision-systems

[119] https://www.dataprotectionreport.com/2024/03/singapore-releases-new-guidelines-on-the-use-of-personal-data-in-ai-systems/

[120] https://www.cliffordchance.com/insights/resources/blogs/talking-tech/en/articles/2023/04/the-italian-data-protection-authority-halts-chatgpt-s-data-proce.html; The DPA's order found that there was 'a material risk that ChatGPT would breach the GDPR on a number of grounds: (i) The users of ChatGPT and other data subjects whose data is processed by OpenAI are not provided with a privacy notice (breach of Art. 13 GDPR) (ii) The use of personal data as a training set for the AI software is unlawful due to the absence of an adequate legal basis (breach of Art. 6 GDPR), (iii), The processing is not accurate, in that the information contained in ChatGPT's responses to users' queries is not always correct (breach of Art. 5 GDPR), (iv) Although OpenAI's terms and conditions forbid access to users below the age of 13, OpenAI has not implemented measures to detect the users' age and block access accordingly (breach of Art. 8 GDPR).' https://www.cliffordchance.com/insights/resources/blogs/talking-tech/en/articles/2023/04/the-italian-data-protection-authority-halts-chatgpt-s-data-proce.html The ban was subsequently lifted 4 weeks later after OpenAI 'addressed or clarified' the issues raised by the DPA; however, in January 2024 OpenAI was notified by the Italian DPA that it was again suspected of violating GDPR. https://techcrunch.com/2024/01/29/chatgpt-italy-gdpr-notification/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAABunj1LF2tp3L75m3QTqWHYh8JT3B93g8WaNoYJPjRIM1TyWf1VA46qyu75_Uy02LYimyuiCEc5AAsP59uHbqj4bDPFqXflH1q_fbx-QVOUyxCV-qL83fMGqIv2NzC-9ekzOJ5YkMVKHc0H3eIN8B5tIQ6g46bbEwG5-OGZTuH

[121] https://www.pinsentmasons.com/out-law/news/uber-case-a-reminder-dangers-potentially-discriminatory-ai#:~:text=It%20follows%20a%20case%20in,under%20the%202010%20Equality%20Act

[122] https://www.justice.gov/opa/pr/justice-department-secures-groundbreaking-settlement-agreement-meta-platforms-formerly-known

[123] https://www.doa-law.com/wp-content/uploads/2024/02/Deepfakes-Legal-Safeguards-in-Nigeria.pdf

ABNASIA.ORG

Existing Legal Framework	Examples of Application to AI Ecosystem
Intellectual property	In late 2023 the New York Times (NYT) sued OpenAI in US courts, arguing that OpenAI had engaged in large-scale copyright infringement, on the grounds that (a) OpenAI's platform is trained on large volumes of the NYT's articles, which are protected by copyright, (b) the LLMS that have been trained are a derivative work of the NYT's body of copyrighted work, and (c) ChatGPT outputs closely mimic NYT articles, in effect reproducing copyrighted material.[124] OpenAI has defended itself on the basis that its use of NYT articles is protected under the 'fair use' doctrine.[125]
Competition / antitrust	The US Federal Trade Commission (FTC) announced in January 2024 that it issued orders to five AI developers (Google, Amazon, Anthropic, Microsoft and OpenAI) requiring them to provide information regarding recent investments and partnerships involving generative AI companies and major cloud service providers, to understand their impact on the competitive landscape.[126]
Procurement	The 2023 Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence seeks to leverage the US government's federal procurement power to set industry standards for AI safety.[127] The Executive Order directs the Office of Management and Budget (OMB) to issue guidance to federal agencies on managing AI risks in the federal government.[128] In September 2024, the OMB issued guidance on responsible AI acquisition by the federal government, setting out three strategic goals: 'managing AI risks and performance,' 'promoting a competitive AI market with innovative acquisition,' and 'ensuring collaboration across the federal government'.

[124] https://hls.harvard.edu/today/does-chatgpt-violate-new-york-times-copyrights/

[125] Id.

[126] https://www.ftc.gov/news-events/news/press-releases/2024/01/ftc-launches-inquiry-generative-ai-investments-partnerships

[127] 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/; https://www.ey.com/en_us/insights/public-policy/key-takeaways-from-the-biden-administration-executive-order-on-ai

[128] Id.

ABNASIA.ORG

The benefit of this approach is that it leverages the existing enforcement expertise, infrastructure, and resources of current supervisory bodies, without the need to pass fresh legislation. These frameworks can be updated under a 'wait and see' approach to regulation – for example, the UK government's position is that, although new legislative action regarding general purpose AI systems will be necessary, doing so at present would be premature. The UK's preferred approach is to empower existing regulatory authorities and frameworks to apply the UK's AI principles (see above) to tackle AI risks.[129] New guidance issued by existing regulators can help industry actors update their existing compliance processes to address AI risks.

However, reliance on existing legal frameworks is constrained by the existing scope of these frameworks. For example, data protection laws often only apply to personal data, and regulators lack the power to regulate AI models that are trained mainly on non-personal data. Because enforcement and supervisory experience relating to AI is scarce, it is unclear how effective the application of some of the above legal frameworks will be – for example, the outcome of the US copyright litigation noted above is highly unclear, meaning the effectiveness of IP and copyright regimes for safeguarding the interests of publishers, content producers, and the creative industry is in question.

Over-reliance on existing legal frameworks can create a patchwork approach to regulation, with potential gaps in rights protection or risk mitigation – this may eventually lead to unacceptable harms for consumers and a lack of legal certainty for industry actors. If existing legal frameworks are leveraged under a 'wait and see' approach, it is critical for policymakers to introduce a monitoring layer within the existing regulatory architecture, to allow early identification of potential gaps in the complex, overlapping legal architecture and ensure that such gaps are addressed in any new legislation or other administrative measures.

Additionally, in EMDE contexts, many of these legal frameworks may not yet exist; where they are in place, they may face gaps (both in terms of the substantive legal framework or in the institutional capacity for regulatory enforcement). In addition, the interaction between existing legal regimes and new laws on AI will be complex. **Questions regarding how countries should prioritize allocation of their policymaking and institutional resources are not well settled and will greatly depend on local exigencies – it will be the task of country policymakers to consult with all relevant stakeholders on the best way forward for local consumers and priorities.**

Technology-specific and sectoral approaches

Legal frameworks governing AI can also be targeted towards certain sectors and product areas or towards specific use cases and application types. This method contrasts with broad, overarching regulations, allowing for more precise control and management of AI technologies in areas where they have unique impacts and risks.

Some non-exhaustive examples of sectoral AI regulation include:

1. **Healthcare:** In the US, the Food and Drug Administration (FDA) is considering updating its existing pre-market review processes to regulate AI and machine learning-enabled medical devices.[130] In India, the Medical Council of India and the Ministry of Health and Family Welfare oversee regulations to ensure AI applications in healthcare comply with data privacy and safety standards.[131] For a deeper discussion of sectoral governance considerations for healthcare, see box 16 below.

[129] https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#executive-summary; https://www2.deloitte.com/uk/en/blog/emea-centre-for-regulatory-strategy/2024/the-uks-framework-for-ai-regulation.html

[130] https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device#regulation

[131] https://www.niti.gov.in/sites/default/files/2021-09/ndhm_strategy_overview.pdf#:~:text=URL%3A%20https%3A%2F%2Fwww.niti.gov.in%2Fsites%2Fdefault%2Ffiles%2F2021

ABNASIA.ORG

2. **Financial Services:** The financial sector employs AI for credit scoring, fraud detection, and algorithmic trading. In July 2023, the US Securities and Exchange Commission (SEC) proposed new rules designed to regulate potential conflicts of interest associated with private funds' use of AI-related technologies in their interactions with investors.[132] The UK Financial Conduct Authority is set to regulate 'critical third parties' that provide critical technologies, including AI, to regulated financial entities.[133]

3. **Transportation:** Autonomous vehicles and AI-driven traffic management systems are regulated to ensure safety and efficiency. Singapore's Land Transport Authority (LTA) has established guidelines for the testing and deployment of autonomous vehicles.[134] The country also implemented a 5-year regulatory sandbox in 2017 to facilitate the safe development and integration of autonomous vehicles.[135]

4. **Employment:** AI systems used in hiring and workplace management are regulated to prevent discrimination and ensure fairness. New York City's Local Law 144[136] requires independent bias audits for automated employment decision tools (see box 17).

[132] https://www.dechert.com/knowledge/onpoint/2023/9/sec-proposes-new-regulatory-framework-for-use-of-ai-by-broker-de.html

[133] https://www.proskauer.com/blog/a-tale-of-two-regulators-the-sec-and-fca-address-ai-regulation-for-private-funds

[134] https://cms.law/en/int/expert-guides/cms-expert-guide-to-autonomous-vehicles-avs/singapore

[135] https://www.ippapublicpolicy.org/file/paper/5cea683b9a45b.pdf

[136] https://www.nyc.gov/site/dca/about/automated-employment-decision-tools.page

ABNASIA.ORG

Box 16: Sectoral Governance and Regulatory Aspects: Health Sector Case Study

Regulating AI and GenAI in healthcare presents complex challenges due to the high stakes involved in patient care and medical decision making. AI systems in healthcare, need to be highly accurate and reliable, as incorrect decisions or predictions could result in misdiagnoses or inappropriate treatments. Moreover, the integration of AI into healthcare workflows raises questions about accountability and the role of healthcare professionals in AI-driven decisions.

In response to these challenges, in January 2024, the World Health Organization (WHO) developed health sector guidelines on the use of AI in healthcare. These guidelines highlight the need to harness AI's benefits while minimizing potential risks. Key regulatory considerations include ensuring the safety and effectiveness of AI systems, implementing strong privacy and security measures, and promoting transparency and trust in AI technologies. WHO also stresses the ethical use of AI, prioritizing human rights, safety, and preventing biases or misinformation that could cause harm. The guidelines provide support for governments, developers, healthcare providers, and other stakeholders in managing AI responsibly in healthcare.

The responsibility for regulating medical devices and medical products lies with countries' medical regulatory authorities, such as the Food and Drug Agency (FDA) in the United States, the Thai Food and Drug Administration in Thailand, the European Medicines Agency (EMA), or the United Kingdom's Medicines and Healthcare products Regulatory Agency (MHRA) among others. These agencies are responsible for ensuring the safety, efficacy, and quality of medical products, including drugs, medical devices, and, increasingly, some types of software used in healthcare (software as a medical device, software in a medical device, and software as an accessory to a medical device). These agencies have already cleared several uses of AI in medical devices; the FDA, for example, has cleared 950 such medical devices as of August 2024, but none of them have included GenAI.

Unlike traditional medical devices, AI models can change post-deployment, making it difficult for current approval frameworks to ensure long-term performance and safety. The US FDA has recognized these challenges and has suggested that post-market evaluation might become a responsibility that falls, at least in part, on healthcare providers and institutions using AI tools. While this allows for continuous oversight as AI evolves, it raises concerns about the burden on providers, who may lack the expertise or resources to effectively monitor AI performance. Additionally, it could lead to inconsistencies in how AI is assessed across different healthcare settings.

Bottom line: Regulating GenAI in healthcare requires a balance between caution and flexibility. Policymakers must integrate sector-specific regulations, maintain and amend existing processes as needed, and evaluate new technologies carefully. At the same time, sector-specific regulatory frameworks may need updating to accommodate the evolving nature and broader applications of technologies like GenAI, ensuring appropriate value, safety, and cost-benefit measures are in place.

Sources: https://www.who.int/publications/i/item/9789240084759; https://jamanetwork.com/journals/jama/fullarticle/2825146

ABNASIA.ORG

Box 17: Regulatory Experience with Algorithmic Auditing – New York City's Local Law 144

Algorithmic auditing tests AI products for risks like discrimination and toxic content. In late 2023, researchers from the Ada Lovelace Institute and Data & Society analyzed New York City's Local Law 144 (LL 144), which mandates independent bias audits for employers using automated decision-making tools. The study found flaws in the law's framework, such as a lack of a robust third-party auditing ecosystem, insufficient requirements to stop using biased tools, and weak enforcement mechanisms. Auditors also faced challenges in accessing necessary data and a lack of standardized practices.

The authors of the report offered 6 recommendations for policymakers designing future algorithmic auditing regimes:

1. Auditing laws must establish clear definitions that capture the full range of AI systems in scope, in consultation with affected communities
2. Auditing laws must establish clear standards of practice on the role and responsibilities of auditors
3. Auditing laws must enable smooth data collection for auditors, including clear procedures and requirements around data access (e.g. which information and documentation about datasets needs to be turned over)
4. Auditing laws must establish meaningful metrics that accurately capture algorithmic risks
5. Audits should follow a theory of change that results in meaningful outcomes and accountability
6. Auditing laws need mechanisms to monitor and enforce against non-compliance.

Source: https://www.adalovelaceinstitute.org/report/code-conduct-ai/

Application-specific regulation targets particular kinds of AI products, such as GenAI applications. China was one of the first jurisdictions to introduce any form of binding legislation governing AI applications – it currently has separate laws regulating recommendation algorithms, 'deep synthesis' technologies (a subset of generative AI technologies that includes deepfakes and digital simulation models), and generative AI services.[137] For a deeper discussion of China's regulatory regime, see Annex 2.

New binding frameworks can also be targeted towards specific elements within the AI stack, such as hardware or other infrastructure. These rules can often be imposed via secondary legislation or other executive action. For example, the US Department of Commerce, Bureau of Industry Security (BIS) has introduced a range of export control measures aimed at restricting the export of advanced semiconductors and other related equipment to China and other countries.[138]

Sector-specific and technology-specific approaches can provide a highly contextual form of elaborating regulation. They can be particularly effective if enforced by sectoral regulators with specific expertise in working with regulated entities. However, as noted above, policymakers should avoid an excessively

[137] https://www.lw.com/admin/upload/SiteAttachments/Chinas-New-AI-Regulations.pdf

[138] https://www.nortonrosefulbright.com/en/knowledge/publications/5a936192/us-expands-export-restrictions-on-advanced-semiconductors

ABNASIA.ORG

fragmented legal regime governing AI, to maximize legal certainty and minimize gaps in protections for consumers regarding AI harms. Laws which are scoped to only apply to certain types of AI systems (e.g. generative AI systems) risk becoming out of date if technological developments create new forms of AI harm that do not neatly map onto existing AI taxonomies.

Audits are crucial for ensuring compliance with established standards and identifying potential risks in AI systems. Regulators can and should leverage audits as a powerful tool to enforce accountability and transparency in AI development and deployment. By regularly assessing AI systems, audits can uncover vulnerabilities and ensure that ethical guidelines are being followed.

Box 18: **President Biden's Executive Order:**

Signed in October 2023, this binding directive directs multiple federal agencies to develop and implement guidelines, standards, and best practices for the safe, secure, and trustworthy development and use of AI technologies. The actions mandated by the executive order include both voluntary guidelines and mandatory requirements for AI developers and organizations, particularly for those creating high-risk AI systems.

Key Roles and Responsibilities The EO underscores the importance of various federal agencies in regulating and securing the development and use of AI. Some key roles and responsibilities highlighted in the order include:

- Evaluation of Misuse Potential: Assigned to the Secretary of Homeland Security, Secretary of Energy, and OSTP, this role involves assessing the potential misuse of AI for developing CBRN (chemical, biological, radiological, or nuclear) threats
- Red-Team Testing Standards: The National Institute of Standards and Technology (NIST) is tasked with setting rigorous standards for red-team testing (see box 23) to ensure the safety of AI systems before public release.
- Content Authentication: The Department of Commerce is responsible for developing guidance on content authentication and watermarking to protect Americans from AI-enabled fraud and deception.
- Military and National Security Oversight: The Department of Defense, Department of State, and the U.S. Intelligence Community are mandated to ensure secure and responsible AI use in military and national security contexts.
- Critical Infrastructure Safeguarding: The Department of Homeland Security (DHS) is assigned the role of safeguarding critical infrastructure against potential AI threats, focusing on resilience and security.

Source: 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/

ABNASIA.ORG

Tool 4: Regulatory Sandboxes

Regulatory sandboxes offer a controlled and time-bound environment for the development and testing of new products and technologies. Sandboxes have been widely adopted in the financial sector, where they first rose to prominence. They enable government agencies to maintain oversight and control while creating a dynamic regulatory environment for testing emerging technologies and business models, thereby generating empirical evidence to inform policy.[139] Regulatory sandboxes enable experimental innovation within a framework of controlled risks, improving regulators' understanding of new technologies.[140]

They can be designed for a range of different policy objectives, meaning that policymakers should first consider their objectives and the problems they are trying to solve before setting up a sandbox, as these choices will define the design and measurement of sandbox outcomes.[141] Most sandboxes share the following features: : (i) they are temporary, (ii) they use an agile, fail-fast and feedback loop approach, and (iii) they involve collaboration and iteration between stakeholders - specifically industry and policymakers.[142] Based on practice from more mature sectors such as Fintech, policymakers globally have begun to pilot regulatory sandboxes for AI[143] – case studies from Colombia, Singapore and Brazil are discussed in Box 19.

[139] Global Experiences from Regulatory Sandboxes, Appaya et al.(2020)

[140] https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf

[141] https://documents1.worldbank.org/curated/en/579101587660589857/pdf/How-Regulators-Respond-To-FinTech-Evaluating-the-Different-Approaches-Sandboxes-and-Beyond.pdf, p.19. In the FinTech context, four sandbox types have been identified: (1) policy focused, seeking to remove regulatory barriers to innovation, identifying if the regulatory framework is fit for purpose, (2) innovation focused, seeking to increase competition in the marketplace and encourage innovation, (3) thematic, focusing on precise policy objectives or supporting developments of particular sub-sectors or products, and (4) cross-border, seeking to support cross-border operation of firms while encouraging regulatory cooperation and harmonization; id at pp. 20-22.

[142] https://www.oecd-ilibrary.org/docserver/8f80a0e6-en.pdf?expires=1715608170&id=id&accname=ocid195787&checksum=B31A4DD15AD2C4539A32D1D779C4372E, p.8

[143] See discussion in https://scholarlycommons.law.cwsl.edu/cwilj/vol54/iss2/3/

ABNASIA.ORG

Box 19: AI Regulatory Sandbox Case Studies: Colombia, Brazil and the EU[144]

Colombia

In collaboration with the Superintendence for Industry and Commerce, Colombia's data protection authority has launched a regulatory sandbox on 'privacy by design' and by default in AI projects.[145] The purpose of this sandbox is to create a controlled environment for AI developers to collaborate with relevant regulatory authorities to develop their products in a manner that is compliant with regulation.[146]

Brazil

In October 2023, Brazil's data protection authority (the ANPD) launched a regulatory sandbox pilot program for AI and data protection – this included a public consultation process with the public and private sectors.[147] Although the impact of the Brazilian regulatory sandbox on the local innovation ecosystem and regulatory compliance is not yet clear due to the early stage of implementation, it takes a broad multi-stakeholder approach, coordinating action between regulators, regulated entities, technology companies, academics and civil society organizations.[148]

EU

In the EU, regulatory sandboxes are planned under the remit of the recently established EU Office. They are designed to foster innovation, particularly for SMEs, by facilitating the training, testing, and validation of AI systems before market entry. The Office will provide technical support, advice, and tools for establishing and operating these sandboxes, coordinating with national authorities to encourage cooperation among Member States. It is expected that two years after the AI Act comes into force, each Member State must establish at least one AI regulatory sandbox, either independently or by joining with other Member States. This supervision aims to provide legal clarity, improve regulatory expertise and policy learning, and enable market access. The AI Office should be notified by national authorities of any suspension in sandbox testing due to significant risks, and national authorities must submit annual reports to the AI Office and AI Board detailing sandbox progress, incidents, and recommendations.[149]

Source: Adapted from Barsalay et al. (2024); https://scholarlycommons.law.cwsl.edu/cwilj/vol54/iss2/3/

[144] Adapted from https://scholarlycommons.law.cwsl.edu/cwilj/vol54/iss2/3/

[145] See SIC announces privacy-by-design and-default sandbox, https://iapp.org/news/a/colombian-dpa-announces-privacy-by-design-and-default-sandbox/

[146] The objectives of this regulatory sandbox are to (i) establish criteria to facilitate compliance with the regulation on data processing in artificial intelligence projects; (ii) ensure that personal data processing is done appropriately; (iii) promote rights-respecting AI products by design; (iv) accompany and advise companies to mitigate associated risks; (v) consolidate a proactive approach towards compliance with human rights in AI projects and (vi) suggest or recommend adjustments, corrections or adaptations to Columbia's regulatory framework for technological advances. See https://www.redipd.org/en/news/colombia-data-protection-authority-launches-innovative-regulatory-sandbox-privacy-design-and. For a more general account of trends in AI Regulation across Latin America, see https://www.accessnow.org/wp-content/uploads/2024/02/LAC-Reporte-regional-de-politicas-de-regulacion-a-la-IA.pdf

[147] ANPD's Call for Contributions to the regulatory sandbox for artificial intelligence and data protection in Brazil is now open, https://www.gov.br/anpd/pt-br/assuntos/noticias/anpds-call-for-contributions-to-the-regulatory-sandbox-for-artificial-intelligence-and-data-protection-in-brazil-is-now-open

[148] https://www.dataguidance.com/news/brazil-anpd-opens-ai-regulation-sandbox-public

[149] https://artificialintelligenceact.eu/the-ai-office-summary/

ABNASIA.ORG

When implemented effectively, regulatory sandboxes can complement traditional regulatory approaches by generating concrete evidence on how certain governance tools interact with AI systems in practice. This allows policymakers to test and evaluate new regulatory methods, ensuring frameworks achieve intended policy objectives while avoiding unintended consequences. Regulatory sandboxes can be combined with other regulatory tools set out in this section, as part of an iterative, evidence-based approach to regulation. Sandboxes can also leverage the expertise of existing supervisory authorities (e.g. data protection authorities) to ensure coordination between different regulatory frameworks (see the discussion below regarding the application of existing bodies of law to AI).[150] A collaborative form of regulation where the oversight authority acts as a partner and not simply an enforcer may be particularly useful in economies where the AI ecosystem is relatively young, given that AI model providers and deployers may not be equipped to comply with stringent legal obligations without hands-on guidance from regulators.[151]

However, it is important to note that sandboxes on their own are not turnkey solutions for AI governance – sandboxes are most useful where there are regulatory questions that can be solved with evidence derived from experimentation.[152] In other circumstances, the resources needed to run a sandbox may outweigh the upsides – a 2020 World Bank study on Fintech sandboxes found that running such sandboxes is extremely resource-intensive and can place great burdens on regulators, diverting resources and limited capacity away from other critical functions.[153] Regulatory sandboxes can also create potential market distortions and unfair competition, as participants in the sandbox have the first mover advantage and may be seen to have the regulator's 'stamp of approval'. Ultimately, sandboxes should not be a substitute for building effective, permanent regulatory and legal frameworks for AI.

Box 19: Case Study: Singapore's AI Verify

Singapore's AI Governance testing framework and toolkit, 'AI Verify,' launched as a pilot in May 2022, is a unique example of a 'light-touch approach' to AI governance and regulation. [154] AI verify validates the performance of AI systems against a set of internationally recognized principles and frameworks through standardized tests. It provides a testing report that serves to inform users, developers, and researchers. The Future of Privacy Forum notes that, 'rather than defining ethical standards, AI Verify provides verifiability by allowing AI system developers and owners to demonstrate their claims about the performance of their AI systems.'[155] The Singaporean approach is notable because it aims to facilitate interoperability with other regulatory frameworks.

Source: https://aiverifyfoundation.sg/

[150] https://scholarlycommons.law.cwsl.edu/cwilj/vol54/iss2/3/

[151] Id.

[152] https://documents1.worldbank.org/curated/en/579101587660589857/pdf/How-Regulators-Respond-To-FinTech-Evaluating-the-Different-Approaches-Sandboxes-and-Beyond.pdf, p.25

[153] https://documents1.worldbank.org/curated/en/912001605241080935/pdf/Global-Experiences-from-Regulatory-Sandboxes.pdf

[154] https://aiverifyfoundation.sg/

[155] https://fpf.org/blog/ai-verify-singapores-ai-governance-testing-initiative-explained/

ABNASIA.ORG

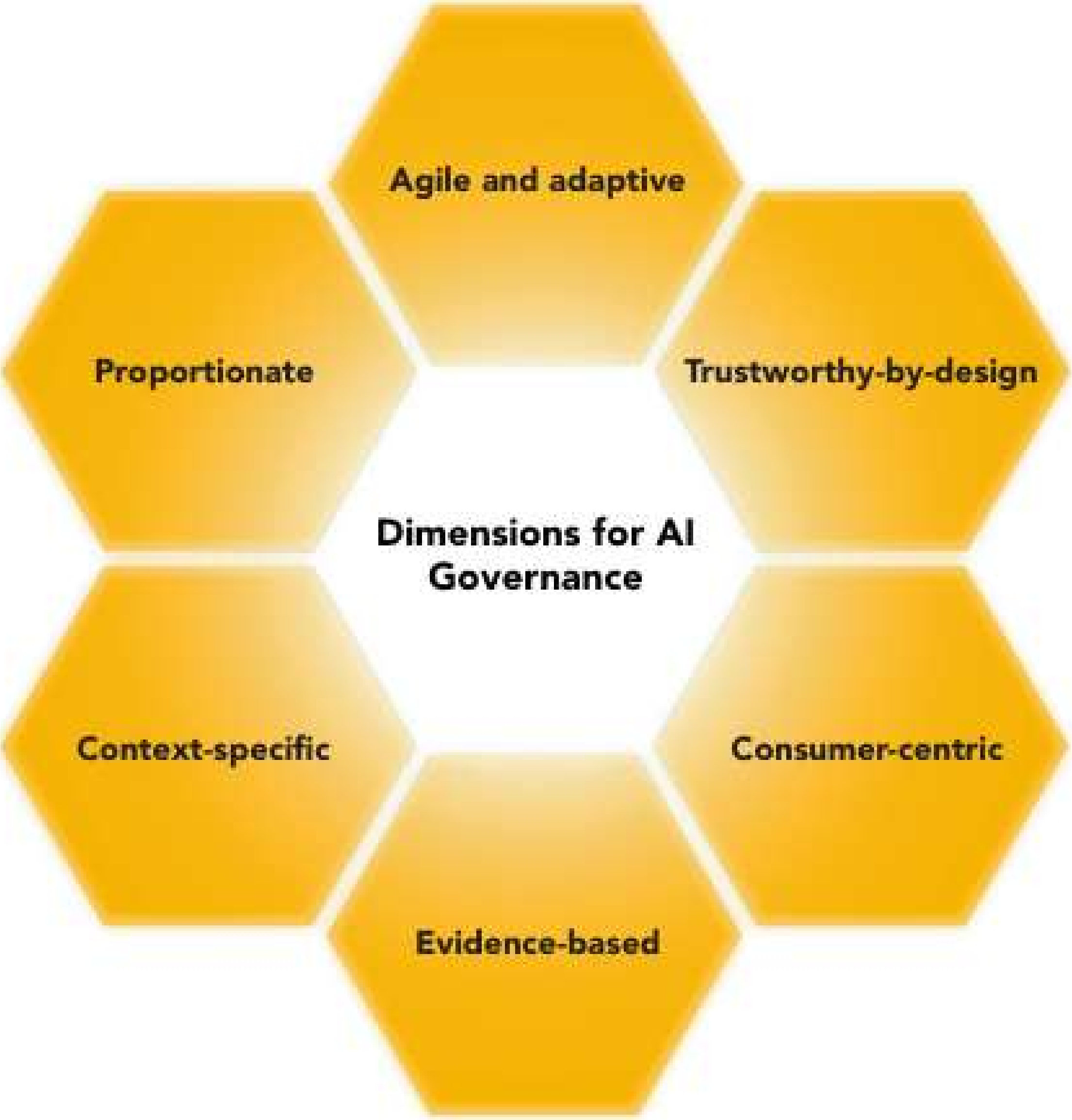
Section 5

Dimensions for AI Governance

ABNASIA.ORG

This section sets out some guiding factors for policymakers when designing their AI governance interventions. These are designed to be resilient to technological, economic and societal changes.

This report puts forward the following 6 preliminary dimensions for designing AI governance frameworks:[156]



Dimensions for AI Governance

Agile and adaptive

Proportionate

Trustworthy-by-design

Context-specific

Consumer-centric

Evidence-based

Figure 4. Preliminary dimensions for designing AI governance frameworks

[156] These principles are intended to guide countries in designing their governance frameworks. However, they share many similarities with substantive principles designed to guide the development of AI systems, such as those set out by the OECD.

ABNASIA.ORG

Table 4.

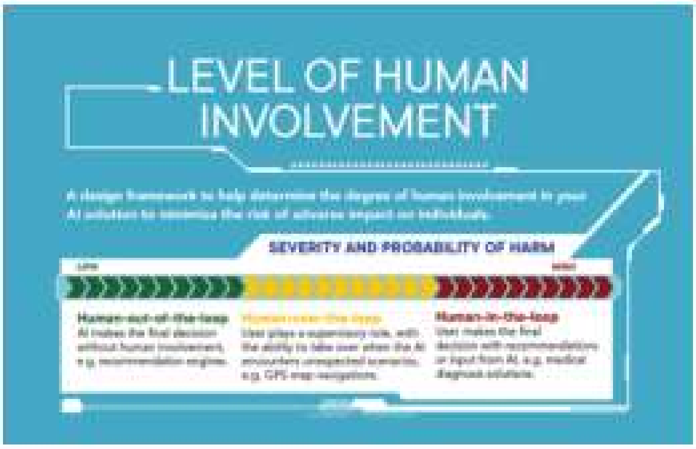
Dimension	Application
Proportionate	**The level and intensity of precautionary requirements can be matched to the risk or scale of the activities being regulated.** The goal is to ensure that governance interventions are effective in managing risks and achieving policy objectives without imposing unnecessary burdens, particularly on smaller or less risky entities. **Principles** Key principles when adopting this approach include: • Risk-Based Approach: Regulations are matched to the risk level, with higher-risk activities facing stricter requirements and lower-risk activities lighter regulations. • Scalability: The regulatory framework adjusts to the size and capacity of entities, reducing the burden on smaller, less risky organizations. • Flexibility: Allows for adjustments over time to keep regulations relevant and effective. **Challenges** Assessment Accuracy: While practical, a proportionate approach also requires a means of assessing risk, harm, and societal impacts, which can be unpredictable or unforeseeable.[157] As a result, some regulators, like those in the United States and Europe rely on proxy indicators, such as the amount of computing power required to train models, the number of parameters or other technical features - which can become out of data with technological advancements.[158] Consistency: Ensuring consistency in regulatory application across different sectors and entities is important to avoid perceptions of unfair treatment. The size of models may also not accurately map onto either the likelihood or severity of harms for affected populations. Moreover, the need for continuous **dynamic adjustments** to keep pace with changes in the industry, technology, and risk landscape can be challenging.

[157] https://arxiv.org/abs/2403.13793

[158] The October 30, 2023 Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence places requirements on AI models trained on 10^26 floating point operations (FLOPs). In addition, it places reporting requirements on AI models of 10^23 FLOPS that use biological sequence data. The February 2024 EU AI Act version relies on 10^25 FLOPs. These indicators may become out of date with advances in computational efficiency requiring less FLOPs to train the same or more powerful models. 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/

ABNASIA.ORG

Dimension	Application
Trustworthy-by-design	Governance frameworks can be designed to encourage AI development that is 'trustworthy-by-design' – embedding trustworthiness throughout the AI lifecycle, from initial conceptualization and data collection through model training, testing, deployment, and monitoring. Relying solely on ex-post regulatory enforcement, or practices like red-teaming during deployment, is not wholly sufficient.[159] **Principles** Proactive governance requires encouraging sociotechnical practices such as stakeholder engagement, governance boards, ethical reviews, and impact assessments, which can be integrated early in the development lifecycle. Human in the loop. Given the risks of AI systems, it is crucial for governance frameworks to encourage AI developers and deployers to maintain a human 'in the loop', ensuring that AI-driven decisions are validated and aligned with human judgment. The effectiveness of AI systems heavily relies on the quality of data, human capital, and the expertise of the interdisciplinary team responsible for their development and deployment.[160] The framework developed by the Government of Singapore can be helpful in this regard (Figure 5). **Challenges** Creating governance frameworks that encourage ex ante trustworthiness requires significant commitment by private sector AI stakeholders and often demands significant upfront investment. However, it avoids costly fixes (and potential regulatory sanctions) down the line. Addressing these challenges requires continuous collaboration among developers, policymakers, and stakeholders.



A design framework to help determine the degree of human involvement in your AI solution to minimise the risk of adverse impact on individuals.

AI makes the final decision without human involvement, e.g. recommendation engines.

User plays a supervisory role, with the ability to take over when the AI encounters unexpected scenarios, e.g. GPS map navigations.

User makes the final decision with recommendations or input from AI, e.g. medical diagnosis solutions.

Figure 5. Level of human involvement in AI deployment

Source: https://file.go.gov.sg/ai-gov-use-cases-2.pdf

[159] 'Institutional Design Principles for Global AI Governance in the Age of Foundation Models and Generative AI,' Safety and Global Governance of Generative AI Report, WFEO-CEIT, Shenzhen Association for Science and Technology, https://www.wfeo.org/wp-content/uploads/2024/CEIT_Safety-and-Global-Governance-of-Generative-AI.pdf

[160] Stankovich (2021)

ABNASIA.ORG

Dimension	Application
Human-centric	A human-centric approach to AI governance places 'the needs and values of people and communities at the center of AI governance and deployment'.[161] Procedurally, this means that viewpoints and inputs from individuals with different backgrounds, interests, and values, as well as the expectations of affected or vulnerable communities is highlighted through the policymaking process.[162] Human-centric AI governance requires designing rules that place fundamental rights and consumer interests as a priority. Human-centric AI governance can incorporate concepts such as data stewardship, that emphasize practices empowering people to inform, shape, and govern their own data.[163] It can also incorporate 'civic tech' tools that help operationalize large-scale engagement and participation in decision-making processes. For example, the 'vTaiwan' project is an 'an open consultation process that brings the Taiwanese consumers and the Taiwanese government together to craft country-wide digital legislation' using collaborative, open-source engagement tools – in 2018 it was reported that 26 issues had been discussed through this open consultation process, with more than 80% leading to government action.[164] **Principles** ▪ User Involvement and Inclusivity: Actively engaging end-users in the policy design and development process to ensure that AI governance frameworks meets their needs and expectations. ▪ Transparency: Ensure that the AI policymaking process is understandable and transparent, providing clear insights into how decisions are made. ▪ Responsiveness: Implement mechanisms for ongoing user feedback and continuously improve governance systems based on this feedback. **Challenges** Implementing a human-centric approach to AI governance presents several challenges. Ensuring broad and meaningful engagement from diverse user groups can be difficult, requiring significant resources and commitment. Additionally, balancing the needs and values of different stakeholders, especially in global contexts, can lead to conflicts and complexities.

[161] https://www.frontiersin.org/articles/10.3389/frai.2023.976887/full

[162] Id.

[163] https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/

[164] https://www.frontiersin.org/articles/10.3389/frai.2023.976887/full

ABNASIA.ORG

Dimension	Application
Agile and adaptive	Agile and adaptive AI regulatory frameworks are designed to be flexible and responsive, allowing for iterative development and rapid adjustments in response to technological and market changes.[165] These frameworks rely on trial and error, co-designing governance frameworks and standards with stakeholders, and incorporating shorter feedback loops. **Principles** • Iterative Approach: Governance frameworks are developed and refined through continuous feedback and iteration, enabling quick responses to changes in technology and market conditions. • Multi-Stakeholder Collaboration: Effective governance requires collaboration among regulators, industry players, academia, and civil society to ensure diverse perspectives and expertise are integrated. • Data-Driven: Utilizing real-time data flows and open data sources to inform regulatory decisions, enhancing the ability to monitor compliance and adapt governance dynamically. **Challenges** Challenges include complex coordination among multiple stakeholders, which can be resource-intensive and require effective communication mechanisms and ensuring that governance keeps pace with rapid technological advancements while maintaining their effectiveness without over-regulating are critical concerns. Striking a balance between being adaptive and providing a stable regulatory environment can also be challenging.

[165] Adapted from OECD (2021) Recommendation of the Council for Agile Regulatory Governance to Harness Innovation.

ABNASIA.ORG

Dimension	Application
Evidence-based	**This emphasizes the need for empirical evidence to demonstrate the impact of regulatory interventions on AI companies' internal safety, ethics, and security practices.** This approach seeks to go beyond self-monitoring by requiring transparent reporting and accountability measures – in other words, AI companies cannot be allowed to 'assign and mark their own homework'.[166] **Principles** • **Empirical Validation:** AI developers must provide empirical evidence of their safety and security practices, moving beyond opaque internal compliance to transparent, verifiable measures. • **Mandatory Reporting:** Mandatory requirements for reporting incidents and corrective measures, similar to cybersecurity frameworks for data breaches. For example, the EU AI Act requires large AI developers to track, document and report serious incidents and possible corrective measures to the AI Office and relevant national competent authorities without undue delay.[167] • **Continuous Monitoring:** Integrating real-time data flows and open data sources to dynamically monitor compliance and effectiveness of AI systems.[168] For example, in deploying AI in healthcare, regulators might monitor products using publicly available data such as software bugs and error reports, customer feedback and social media. Integrating data flows can allow for automation in the regulatory process. Enforcement becomes dynamic, with review and monitoring built into the system.[169] **Challenges** Balancing the need for transparency with the protection of proprietary information and trade secrets is a significant challenge in evidence-based AI governance. Additionally, ensuring that evidence-based measures evolve alongside advancements in AI capabilities and risk mitigation strategies is crucial to maintaining effective and relevant governance interventions.

[166] https://ainowinstitute.org/general/ai-now-joins-civil-society-groups-in-statement-calling-for-regulation-to-protect-the-public

[167] https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

[168] International Telecommunication Union (2020).

[169] World Bank Digital Regulation Platform (2021).

ABNASIA.ORG

Dimension	Application
Context-specific	Context-specific AI governance involves evaluating AI risks and impacts within the specific environments in which AI systems are deployed.[170] AI safety is not an inherent model property; instead, AI safety questions depend 'to a large extent on the context and the environment in which the AI model or AI system is deployed.'[171] Unlike conventional goods that are tested for product safety (e.g. drugs, airplanes, cars), it is difficult to specify in advance how general-purpose AI models will be used and the environments in which they will be embedded – for example, 'evaluations of a foundation model like GPT-4 tell us very little about the overall safety of a product built on it (for example, an app built on GPT-4).'[172] **Principles** • Contextual Evaluation: Assessing AI risks and impacts in the specific deployment environment, considering how the AI system will interact with real human behaviors and societal norms.[173] • Sector-Specific Focus: Supporting sector-specific regulatory enforcement and civil society action to address unique challenges and requirements of different industries. **Challenges** Implementing context-specific AI governance involves accounting for the vast diversity of deployment environments and use cases for AI systems, which can complicate risk assessments and regulatory measures. It requires interdisciplinary expertise to fully understand the context-specific impacts and ensure comprehensive governance. This may require independent public interest research into context-specific AI safety issues, supporting sector-specific regulatory enforcement and civil society action.[174]

[170] Adapted from https://ainowinstitute.org/general/ai-now-joins-civil-society-groups-in-statement-calling-for-regulation-to-protect-the-public

[171] https://www.aisnakeoil.com/p/ai-safety-is-not-a-model-property

[172] https://www.adalovelaceinstitute.org/blog/safety-first/

[173] https://www.trailofbits.com/documents/Toward_comprehensive_risk_assessments.pdf

[174] https://www.adalovelaceinstitute.org/blog/safety-first/

ABNASIA.ORG

Section 6

Stakeholder Ecosystem & Institutional Frameworks

ABNASIA.ORG

Effective design, implementation and supervision of the governance tools described above requires active input and coordination between a wide range of stakeholders. This section aims to map some of the key functions and roles of these stakeholders. While this note does not go into details of all the different institutional arrangements in place for AI - which will be the topic of a forthcoming note - we have highlighted some important governance arrangements here.

The stakeholder ecosystem for AI encompasses a diverse group, including the **public sector** who establish and enforce regulation and policies, **private sector players** who develop and deploy AI technologies, **civil society, academic, and research institutions organizations** that advocate for ethical and equitable AI use and advance AI knowledge, and **end-users** who interact with AI systems and provide valuable feedback. Additionally, the **international community,** including international organizations and multinational partnerships, plays a crucial role in harmonizing standards, fostering cooperation, and addressing cross-border challenges in AI development and deployment. This ecosystem requires collaboration and communication among all parties to ensure AI is developed and used responsibly, ethically, and beneficially.

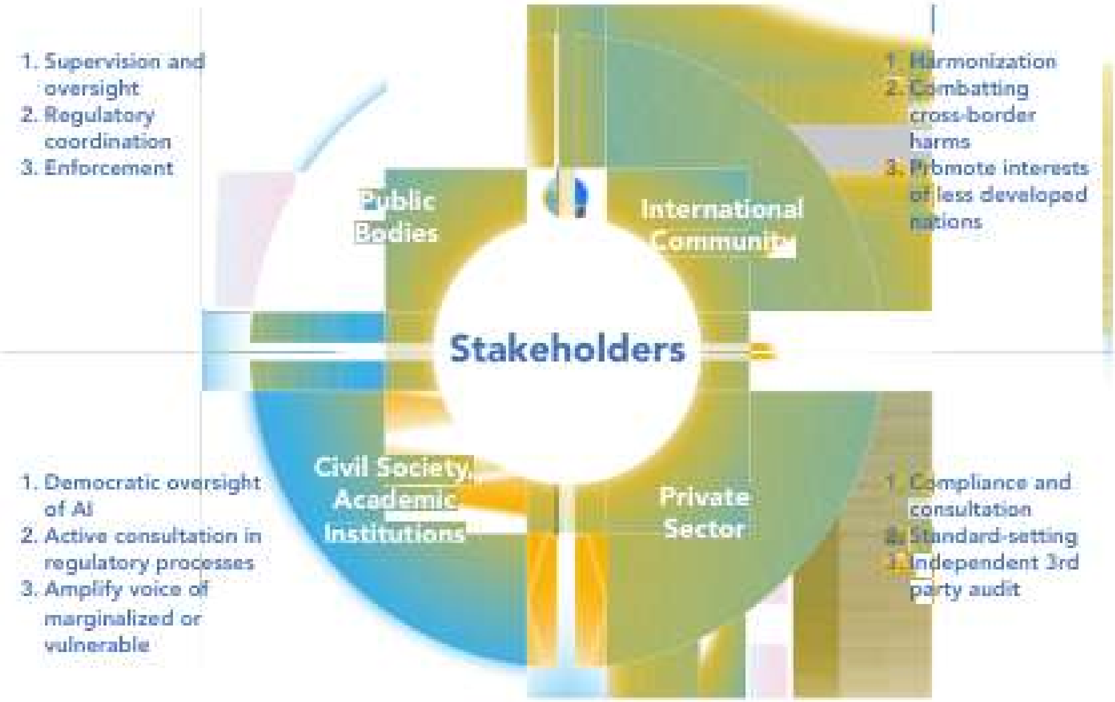

1. Supervision and oversight
2. Regulatory coordination
3. Enforcement
Public Bodies
International Community
1. Harmonization
2. Combatting cross-border harms
3. Promote interests of less developed nations
Stakeholders
1. Democratic oversight of AI
2. Active consultation in regulatory processes
3. Amplify voice of marginalized or vulnerable
Civil Society, Academic Institutions
Private Sector
1. Compliance and consultation
2. Standard-setting
3. Independent 3rd party audit

Figure 6.
Source: Authors

ABNASIA.ORG

6.1. Public and Regulatory Bodies

A number of countries are developing comprehensive AI strategies to guide the national development of AI technologies. These strategies are spearheaded by various ministries depending on the country. For example, in Canada[175], the Ministry of Innovation, Science, and Economic Development (ISED) leads the AI strategy initiatives. In Estonia[176], the Ministry of Economic Affairs and Communications is responsible for digital and AI policies, while in Japan[177], the Ministry of Economy, Trade, and Industry (METI) takes charge of AI strategy development. However, **the enforcement of AI regulations is typically managed by different entities or regulatory bodies.**

Both existing and new national regulatory institutions play key roles in designing and implementing regulatory tools for AI. In some jurisdictions, enforcement of AI regulations is delegated to existing Data Protection Authorities (DPAs) - for instance, the European Data Protection Board (EDPB)[178] has recommended that DPAs be designated as the 'market surveillance authorities' responsible for enforcing obligations for high-risk AI systems under the EU AI Act.[179]

However, some jurisdictions are also establishing dedicated bodies to oversee AI governance. For instance, the European Union has set up the EU AI Office (see box below), tasked with implementing the AI Act, with a particular focus on general-purpose AI systems. Similarly, the United Kingdom has created the Responsible Technology Adoption Unit (RTA) (previously the Centre for Data Ethics and Innovation (CDEI))[180] to advise the government on the responsible use of AI and data-driven technologies.

[175] https://ised-isde.canada.ca/site/ai-strategy/en

[176] https://digital-skills-jobs.europa.eu/en/actions/national-initiatives/national-strategies/estonia-estonian-digital-agenda-2030

[177] https://www.meti.go.jp/english/press/2022/0128_003.html

[178] https://www.edpb.europa.eu/our-work-tools/our-documents/statements/statement-32024-data-protection-authorities-role-artificial_en

[179] https://www.edpb.europa.eu/news/news/2024/edpb-adopts-statement-dpas-role-ai-act-framework-eu-us-data-privacy-framework-faq_en

[180] https://www.gov.uk/government/news/the-cdei-is-now-the-responsible-technology-adoption-unit

ABNASIA.ORG

Box 20: Implementation and enforcement of the EU AI Act

Enforcement of the EU AI Act will involve a combination of national and supranational authorities.

National Notifying and Market Surveillance Authorities

At the national level, EU member states will establish one notifying authority and one market surveillance authority and must ensure that these national competent authorities have adequate technical, financial and human resources, and infrastructure to fulfil their tasks under the Act.[181] The market surveillance authority is the primary body responsible for enforcement at the national level, and will report to the Commission and relevant national competition authorities on an annual basis.[182]

Much of the future complexity in implementing the AI Act is that there are a huge range of design choices that need to be undertaken at national level – see table 5 for more.

The EU AI Office & Board

At the supranational level, the EU AI Board and the EU AI Office are two distinct entities within the governance framework of the AI Act, each with specific roles and responsibilities.[183]

The *EU AI Office*, established within the Directorate-General for Communication Networks, Content and Technology (DG CNECT) of the European Commission, will work to ensure consistent application of the AI Act across the EU. It will work directly with providers and will monitor, supervise, and enforce the AI Act requirements across the 27 EU Member States. It also aims to facilitate legal clarity and market access by developing voluntary codes of practice, which create a presumption of conformity for AI model providers. Additionally, the AI Office will lead international cooperation on AI, strengthen ties between the European Commission and the scientific community, and serve as the Secretariat to the EU AI Board.[184]

The *EU AI Board* serves as an advisory and coordinating body, composed of representatives from Member States and other relevant entities. It advises the European Commission on AI-related issues – its primary role is to ensure the harmonization of AI regulations, provide guidance, and resolve disputes among national authorities.[185]

Source: adapted from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4817755

Delegation of regulatory authority for enforcing AI obligations can be designed in several ways, with each having their own benefits and drawbacks.[186]

[181] EU AI Act, Art 70.

[182] Art 74, EU AI Act.

[183] https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2982

[184] https://artificialintelligenceact.eu/the-ai-office-summary/

[185] Ibid.

[186] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4817755

ABNASIA.ORG

Table 5.

Option	Benefits	Drawbacks
Create a new national agency specifically dedicated to enforcement of new AI rules.	Centralized body resourced specifically for AI oversight will have experts with specific AI skills.	New regulatory body will lack industry-specific expertise. Process for setting up a new regulatory institution is costly and time-intensive.
Assign responsibility for enforcing new AI rules to existing sectorial regulator (e.g. banking, telecoms, data protection, competition).	Leverages existing organizational framework and sectoral knowledge. Enables evaluation of AI systems in specific deployment contexts.	Potential for disputes over jurisdiction of each regulator, depending on scope of existing legal powers. Potential for 'path dependency' focus on issues familiar to existing regulators, overlooking novel or emergent AI harms.
Establish a 'competence center' within existing authority with AI experience, e.g. banking or network regulator.	Brings together AI experts from different backgrounds and sectors (on temporary or permanent basis) to form interdisciplinary teams on specific cases.	Potential recruitment challenges, particularly for inter-disciplinary technical positions. Relatively novel approach with lack of established precedent.

Source: adapted from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4817755

Practical experience in implementing these design choices is scarce at present. Regardless of the approach taken, it is important to ensure coordination between new AI regulators and existing regulatory institutions – for example, law firm DLA Piper has noted that there are important areas of overlap in the substantive EU rules governing data protection (in the GPDR) and AI (in the AI Act)[187] – regulators need to coordinate to ensure that regulatory resources are expended in the most efficient manner possible, avoiding both duplicated efforts as well as gaps in regulatory supervision.

Existing regulatory institutions will have a large role to play, even in the absence of a new binding AI law. Even in a de-centralized supervisory model, there may still need to be some central monitoring and coordination functions established by the government, to facilitate coordinated regulatory action. For example, the UK's white paper on AI regulation recognizes that a patchwork approach to regulation with little central coordination or oversight may in fact create barriers to innovation due to a lack of coherence and clarity in regulatory obligations – as such, the UK has committed to create a set of centralized mechanisms to ensure the sectoral approach to AI regulation can be monitored and adapted, as well as facilitate a single point of collaboration for all interested parties (international partners, industry, civil society, academia and the public). For more detail, see figure 7.

[187] https://privacymatters.dlapiper.com/2024/04/europe-the-eu-ai-acts-relationship-with-data-protection-law-key-takeaways/

ABNASIA.ORG

Figure 7. Centralized Risk Function within De-Centralized, Sector-Based Regulatory Approach

Source: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#section323

National AI safety institutes are also emerging as crucial entities dedicated to researching, understanding, and mitigating the risks associated with AI. While each institute has its own distinct policy goals and mandates,[188] they focus on ensuring that AI systems are safe, reliable, and aligned with human values. Importantly, for AI Safety Institutes with a quasi-regulatory mandate (e.g. pre-release safety testing), these bodies need to collaborate with existing sectoral regulators or be accompanied by other regulatory interventions to mandate transparency and auditability, to prevent industry actors from refusing to allow access to models.[189] Some established Safety Institutes include the UK AI Safety Institute,[190] the US AI Safety Institute consortium (under NIST),[191] the Japanese AI Safety Institute (within the Information-technology Promotion Agency),[192] and the Canadian AI Safety Institute[193] among others. The effectiveness of these institutes will depend on the scope of their mandate and their resources.

Several of these Institutes have also organized AI Safety Summits to bring stakeholders together to address the critical challenges and risks posed by advanced AI technologies. They also importantly contribute to public awareness and education on AI safety. The first Safety Summit was held in Bletchley in November 2023 (see box 21), this was then followed by a summit in Seoul in May 2024, with plans for another in Paris planned for 2025.

[188] For a comparison of the mandates and functions of the various AI Safety Institutes, see Forum For Cooperation on AI, Briefing Booklet, Dialogue on Artificial Intelligence #22 (on file with author).

[189] https://www.adalovelaceinstitute.org/blog/safety-first/

[190] https://www.gov.uk/government/publications/ai-safety-institute-overview/introducing-the-ai-safety-institute

[191] https://www.commerce.gov/news/press-releases/2024/02/biden-harris-administration-announces-first-ever-consortium-dedicated

[192] https://aisi.go.jp/

[193] https://www.pm.gc.ca/en/news/news-releases/2024/04/07/securing-canadas-ai

ABNASIA.ORG

Box 21: UK AI Safety Institute and Summit and Pre-Evaluations of Models

The creation of the UK AI Safety Institute, hailed as the first state-backed organization focusing on advancing AI safety for the public interest, was announced at the UK AI Safety Summit in November 2023.

The UK AI Safety Institute has three core functions:[194]

1. Develop and conduct evaluations on advanced AI systems, to assess safety-relevant capabilities, safety and security of systems, and societal impacts.
2. Drive foundational AI safety research, through exploratory research projects and convening external researchers.
3. Facilitate information exchange, by establishing clear, voluntary channels for sharing information with national and international stakeholders, subject to privacy and data regulations.

On the back of the UK AI Safety Summit – which was purported to be the first global AI Summit- the Bletchley Declaration[195] was endorsed by twenty-eight nations. The event was useful in defying previous concerns of intense competition among top AI advanced countries i.e. UK, US, and China. Instead, these nations, along with others like Brazil, India, and Indonesia, pledged to engage in collaborative AI safety research, emphasizing the implementation of safety tests before the release of new products. The signatory countries expressed a shared commitment to international cooperation, aiming to drive inclusive economic growth, sustainable development, innovation, protect human rights, and instill public trust in AI systems.

Further, during the UK AI Safety Summit, 8 leading AI tech companies'[196] voluntarily committed to subject their models to pre-release safety testing. However, in April 2024 it was revealed that, although the UK government has said it has begun pre-deployment testing, the AI Safety Institute has only been able to gain access to models after release in most cases (with only London-headquartered Google DeepMind offering a form of pre-deployment access to its Gemini models).[197] Although both OpenAI and Meta were set to imminently roll out their next-generation models (OpenAI's GPT-5 and Meta's Llama-3), neither company had granted access to the UK AI Safety Institute to conduct pre-release testing.[198]

As such, the ability of these bodies to function effectively as part of a 'wait and see' approach may be conditional on the introduction of mandatory transparency and audit requirements, imposed through hard law or via other regulatory avenues.

The Ada Lovelace Institute has recommended several improvements to the UK AI Safety Institute:[199]

1. Integrate the AI Safety Institute into existing regulatory frameworks by working with sectoral regulators to test AI products in particular contexts for safety and efficacy.
2. Give the AI Safety Institute legal authority to compel companies to provide access to AI models, training data, relevant documentation, and information about the model supply chain (including energy/water costs and labor practices).
3. Give the AI Safety Institute and downstream regulators the power to block release of models that pose safety risks ('pre-market approvals' powers).

Source: https://assets.publishing.service.gov.uk/media/65438d159e05fd0014be7bd9/introducing-ai-safety-institute-web-accessible.pdf; https://www.adalovelaceinstitute.org/blog/safety-first/.

[194] https://assets.publishing.service.gov.uk/media/65438d159e05fd0014be7bd9/introducing-ai-safety-institute-web-accessible.pdf, p.8

[195] https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

[196] Amazon Web Services, Anthropic, Google, Google DeepMind, Inflection AI, Meta, Microsoft, Mistral AI and Open AI.

[197] https://www.politico.eu/article/rishi-sunak-ai-testing-tech-ai-safety-institute/

[198] Id.

[199] https://www.adalovelaceinstitute.org/blog/safety-first/

ABNASIA.ORG

More generally, the public sector has an important role to play in shaping the responsible AI ecosystem, by leveraging the economic weight of the government's purchasing power to instigate changes in how large AI companies design and deliver AI solutions – procurement policy frameworks play a crucial role here. As governments increasingly seek to integrate AI into the provision of public sector services, procurement regulations and guidelines will play an increasingly important role in setting robust guardrails and promoting public trust, by ensuring that any use of AI is effective, proportionate, legitimate and in line with broader public sector duties. This is especially important given that many governments will rely on the expertise of AI providers and will likely be AI 'consumers' rather than AI 'developers'.

Box 22: WEF AI Government Procurement Guidelines (2020)

In June 2020 the World Economic Forum (WEF) released a set of 10 guidelines for AI government procurement, outlining key considerations when starting a procurement process, writing a request for proposal (RFP), and evaluating RFP responses:

1. Use procurement processes that focus not on prescribing a specific solution but rather on outlining problems and opportunities, and allow room for iteration.
2. Define the public benefit of using AI while assessing risks.
3. Align your procurement with relevant existing governmental strategies and contribute to their further improvement.
4. Incorporate potentially relevant legislation and codes of practice in your RFP.
5. Articulate the technical and administrative feasibility of accessing relevant data.
6. Highlight the technical and ethical limitations of intended uses of data to avoid issues such as historical data bias
7. Work with a diverse, multidisciplinary team.
8. Focus throughout the procurement process on mechanisms of algorithmic accountability and of transparency norms.
9. Implement a process for the continued engagement of the AI provider with the acquiring entity for knowledge transfer and long-term risk assessment.
10. Create the conditions for a level and fair playing field among AI solution providers.

Source: https://www3.weforum.org/docs/WEF_AI_Procurement_in_a_Box_AI_Government_Procurement_Guidelines_2020.pdf

ABNASIA.ORG

6.2. Private sector

Because the development of cutting-edge AI models is predominantly led by a few large technology companies, the private sector plays a vital role in ensuring responsible AI practices. Policymakers need to consult with industry players and bodies to develop a robust AI governance roadmap. At the same time, caution must be taken to ensure that AI governance does not become subject to industry capture – ultimately responsibility for AI policy should remain with the state, acting in the interests of all consumers and stakeholders, and should not be inappropriately delegated to private actors.

Private sector involvement in AI governance is crucial for several reasons:

- **Innovation and expertise:** The private sector drives much of the innovation in AI and possesses deep technical expertise that can inform effective governance.
- **Resource availability:** Large technology companies have the resources to conduct thorough testing and validation of AI systems, contributing to safer and more reliable AI deployment.
- **Market implementation:** As the primary developers and deployers of AI technologies, private companies are well-positioned to implement governance frameworks and ensure compliance with regulatory standards.

Drawing on experience from other sectors, some effective interactions include:

Ensuring compliance and building trust: The private sector is integral in ensuring compliance with regulatory requirements and building trust in AI systems. Red-teaming and adversarial testing are critical practices where private sector involvement is essential (see Box 23)

Collaborative governance and standardization: Standardization processes provide an important avenue for the private sector to collaborate with other stakeholders to design AI governance frameworks. Outside of multi-stakeholder standards processes, large AI developers play a critical role in mainstreaming good AI governance practices across the ecosystem, through licensing or other contracting frameworks. This is important given that the private sector AI ecosystem includes not just model developers, but downstream deployers of AI across a range of industry areas, including banking, healthcare, education, retail, transportation, energy, etc.

Third-party oversight and audits: Private sector actors can also play a useful role in the AI audit ecosystem by providing independent third-party oversight and testing of AI systems, to validate their impact and safety in context. Such third-party AI audit firms could play a role in a binding regulatory framework, in the same way that accounting firms audit the books of private companies[200] – in such an ecosystem, robust professional and ethical safeguards would need to be put in place to mediate conflicts of interests.

Public-Private Partnerships: Public-private AI partnership may be useful where governments need specific inputs or expertise from AI practitioners – for example, can provide training to regulatory bodies on the harms posed by the latest AI models and help democratize AI development. For instance, in January 2024, the US National Science Foundation announced the National Artificial Intelligence Research Resource, providing a platform for US AI researchers to access computational, data, and training resources donated by large AI companies.[201] However, the AI Now Institute has cautioned that the incentives of private sector actors within public-private partnerships must be carefully scrutinized AI companies need to articulate a robust vision for how public funds' investment within a public-private structure will advance the public good and benefit society at large, ensuring public funding does not simply enable innovation benefits to accrue to incumbent players.[202]

[200] https://www.anthropic.com/news/third-party-testing

[201] https://foreignpolicy.com/2024/02/12/ai-public-private-partnerships-task-force-nairr/, https://nairrpilot.org/about

[202] https://foreignpolicy.com/2024/02/12/ai-public-private-partnerships-task-force-nairr/

ABNASIA.ORG

Box 23: Red-teaming and adversarial testing

Red-teaming is a critical practice in the AI industry aimed at ensuring the robustness and security of AI systems. It involves a structured testing approach where dedicated teams, known as red teams, use adversarial methods to identify vulnerabilities, flaws, and potential risks in AI models. This practice is essential for uncovering harmful or unintended behaviors that could arise from the deployment of AI systems. For instance, the Biden administration's executive order on AI requires high-risk generative AI models to undergo red-teaming, defined as 'a structured testing effort to find flaws and vulnerabilities in an AI system.'[203]

Key Aspects of Red-Teaming:

1. *Identification of Vulnerabilities:* Red teams simulate attacks on AI systems to identify weaknesses that could be exploited by malicious actors, including biases, discriminatory outputs, and other harmful behaviors not evident during standard testing procedures.
2. *Adversarial Testing:* This involves deliberately attempting to cause the AI system to fail or produce incorrect results. By exposing the system to various adversarial scenarios, red teams can identify potential points of failure and areas requiring strengthening.
3. *Internal vs. External Red Teams:* Companies can choose between internal red teams composed of their employees or external red teams made up of independent experts. Internal teams benefit from a deep understanding of the company's systems, while external teams bring a fresh perspective and can often identify issues that internal teams might overlook. For example, Google uses internal red teams, whereas OpenAI creates a network of external red-teamers.
4. *Customized Approach:* The structure and methodology of red-teaming should be tailored to the specific AI system and its deployment context. High-risk AI systems, such as those used in healthcare or finance, may require more rigorous and comprehensive red-teaming efforts.
5. *Continuous Improvement:* Red-teaming should be an ongoing process. As AI systems evolve and new threats emerge, continuous red-teaming efforts are necessary to ensure the systems remain secure and reliable.

Red-teaming is crucial for ensuring the robustness and security of AI systems. It helps identify and mitigate security vulnerabilities before they can be exploited, enhancing the overall security of AI systems. Additionally, red-teaming supports compliance with regulatory requirements by providing evidence that AI systems have been rigorously tested for safety and reliability. By proactively identifying and addressing potential issues, organizations can build trust with users, stakeholders, and regulators, demonstrating their commitment to responsible AI deployment.

However, red-teaming is not a one-size-fits-all solution. Jiahao Chen, director of AI/ML at the NUC Office of Technology and Innovation, notes that red teaming is an assurance framework ensuring AI systems operate as intended. They are not a substitute for audit frameworks which ensure that private sector entities fulfill their regulatory and ethical responsibilities.[204] A policy brief from Data & Society recommends that red-teaming should be accompanied by full accountability measures such as algorithmic impact assessments, external audits, and public consultation.[205]

Source: https://hbr.org/2024/01/how-to-red-team-a-gen-ai-model https://www.ibm.com/think/topics/red-teaming

[203] 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/, section 3(d).

[204] https://www.linkedin.com/pulse/red-teaming-assurance-accountability-jiahao-chen-pzj9e/

[205] https://datasociety.net/library/ai-red-teaming-is-not-a-one-stop-solution-to-ai-harms-recommendations-for-using-red-teaming-for-ai-accountability/

ABNASIA.ORG

6.3. Civil society and direct public participation

Civil society organizations have a critical role to play in representing the interests of consumers and marginalized groups in policymaking fora.

First, civil society can provide an active layer of democratic, regulatory oversight, holding AI developers and deployers accountable to public interests and goals.[206] In particular, civil society organizations can provide an additional layer of accountability by actively auditing and evaluating AI-driven practices on the ground. This role requires governance interventions that mandate transparency and access to critical data flows. For example, the EU actively supports independent fact-checking organizations such as the European Digital Media Observatory (EDMO) and European Fact-Checking Standards Network (EFSCN) as part of their wider approach to combatting disinformation.[207]

Second, civil society has an important role in amplifying the voice of vulnerable or marginalized populations. This can take the form of feedback on formal legislation and regulatory initiatives, as well as inputs representing consumers' interests and fundamental rights at standards development organizations – although their ability to participate in standard-setting activities is limited at present, this should be expanded.

To facilitate these activities, policymakers should ensure a genuine multi-stakeholder, user-driven approach to crafting AI governance interventions. This requires policy interventions that provide civil society with the necessary civic space, financial, human, and technical resources to conduct this work and being cognizant of the potential barriers to participation.

Ensuring opportunities for direct public input in legislative and regulatory processes is also crucial. Some governments have already begun experimenting with ways of encouraging direct democratic consumer engagement in AI governance – in February 2024 Belgium launched a consumers' panel on AI, comprised of 60 people selected at random bringing together a diverse group in terms of age, gender, education levels, and other demographic criteria.[208] The panel's conclusions were presented to Belgian and European political leaders, as an input to inform Belgium's positions with the Council of the EU when defining the European strategic agenda for 2024-2027.

Academics and research institutions are fundamental to advancing AI knowledge and informing policy development. Their research provides the evidence base needed to understand AI's impacts, risks, and benefits. Academics contribute to setting ethical guidelines and developing best practices for AI deployment. By conducting independent studies and publishing findings, they offer critical insights that help shape effective and responsible AI governance frameworks. Academics also play a key role in educating the next generation of AI practitioners.[209]

[206] https://wiserd.ac.uk/blog/civil-society-perspectives-on-ai-in-the-eu/#:~:text=Namely%2C%20that%20civil%20society%20involvement,founded%20on%20accountability%20and%20transparency.

[207] https://commission.europa.eu/topics/strategic-communication-and-tackling-disinformation/supporting-fact-checking-and-civil-society-organisations_en

[208] https://belgian-presidency.consilium.europa.eu/en/news/launch-of-citizens-panel-on-artificial-intelligence/ See also https://tai.tv/articles/we-need-to-democratize-ai-helene-landemore-john-tasioulas-aud-2680 on the role of citizen assemblies in AI governance.

[209] Id., https://belgian-presidency.consilium.europa.eu/en/news/the-citizens-panel-on-ai-issues-its-report/

ABNASIA.ORG

6.4. International community

International coordination on AI regulation and governance is critical for several reasons.[210]

First, the scale and transboundary nature of AI systems can lead to cross-border impacts and harms. For instance, privacy harms arising from mass data collection are unlikely to be confined to a single country. Similarly, biased outputs can be generated from an AI system initially built in Western Europe but deployed globally, with the potential for harm spreading rapidly across borders and populations. In the absence of global cooperation on rulemaking, large technology companies leading AI development and deployment based overwhelmingly in the Global North may choose only to comply with the regulatory frameworks applicable in their large priority markets.[211] **Smaller EMDEs may therefore find it difficult to exert regulatory influence over AI developers based in other jurisdictions** due imbalances in market size and/or relative geopolitical influence. International coordination on AI governance, if created in a participatory and robust manner, gives EMDEs the opportunity to ensure their needs and concerns are reflected in how AI is governed.

Second, international coordination on AI governance is needed in order to prevent a 'race to the bottom' – i.e. to prevent 'regulatory arbitrage' as private firms seek to relocate their most harmful activities to areas of low regulatory barriers. Ensuring a level regulatory playing field is particularly important for smaller, less developed states who may otherwise face pressures to lower guardrails in order to encourage local innovation or foster foreign investment.

Third, international coordination on AI governance can encourage responsible innovation by lowering compliance costs for businesses – the cost of complying with dozens of fragmented national rules disproportionately disadvantages new startups and market entrants, who do not have the same compliance resources as larger companies.

As countries continue to establish and refine their institutional arrangements, international cooperation remains vital. The Global Partnership on Artificial Intelligence (GPAI), an initiative involving multiple countries, aims to bridge the gap between theory and practice on AI. GPAI facilitates international collaboration by bringing together experts from various fields to promote responsible AI development. Additionally, the OECD has established the AI Policy Observatory, which provides a platform for countries to share best practices and align their AI policies. Furthermore countries are entering into Memoranda of Understanding (MoUs) to facilitate collaboration and harmonize AI policies. For instance the EU and the US have agreed to increase co-operation in the development of technologies based on AI, with an emphasis on safety and governance,[212] while the UK and US Safety Institutes are collaborating to formulate a framework to test the safety of LLMs[213], while at the AI Seoul Summit in May 2024, 10 countries and the EU agreed to launch an international network dedicated to advancing the science of AI safety.[214] Global dialogue is due to continue into 2025, beginning with the AI Action Summit planned for February 2025 hosted by France, which will include a track on global AI governance which aims to shape an effective and inclusive framework for AI governance.

[210] See Veale et al., (2023), p. 265-6.

[211] See Bradford (2019). https://academic-oup-com.libproxy-wb.imf.org/book/36491?login=true&token=

[212] https://www.cio.com/article/2083973/eu-and-us-agree-to-chart-common-course-on-ai-regulation.html

[213] https://www.commerce.gov/news/press-releases/2024/04/us-and-uk-announce-partnership-science-ai-safety

[214] https://www.gov.uk/government/news/global-leaders-agree-to-launch-first-international-network-of-ai-safety-institutes-to-boost-understanding-of-aim

ABNASIA.ORG

Multilaterals and intergovernmental organizations are also increasing their strategic support for AI. The UN is currently engaged in several projects designed to coordinate international AI governance including the Global Digital Compact which aligns countries on a common, inclusive digital development agenda.[215] In parallel, the UN Advisory Body on Artificial Intelligence has outlined plans to establish an inclusive AI governance institution, aiming to harmonize efforts across various global initiatives, building on existing processes. At the World Bank, digitization has been identified as a core Global Challenge Programs to focus funding in the coming years.[216] The World Bank also has an ongoing project, funded by its Human Rights, Inclusion and Empowerment Umbrella Trust Fund, seeking to design AI governance, risk-mitigation and safeguards for Bank-funded projects with AI components. A number of development organizations are also increasingly providing technical assistance and capacity building on AI strategy and policy interventions – and will have an increasingly important role to play in helping their recipient countries craft tailored AI governance frameworks.[217]

[215] https://www.un.org/techenvoy/global-digital-compact

[216] https://www.devcommittee.org/content/dam/sites/devcommittee/doc/documents/2023/Final%20Updated%20Evolution%20Paper%20DC2023-0003.pdf

[217] See Jeremy Ng, 'MDBs and The Legal Fabric of Global AI Governance: Infrastructure as Regulation in the Global Majority' AI In Society (OUP, 2024, forthcoming).

ABNASIA.ORG

Section 7

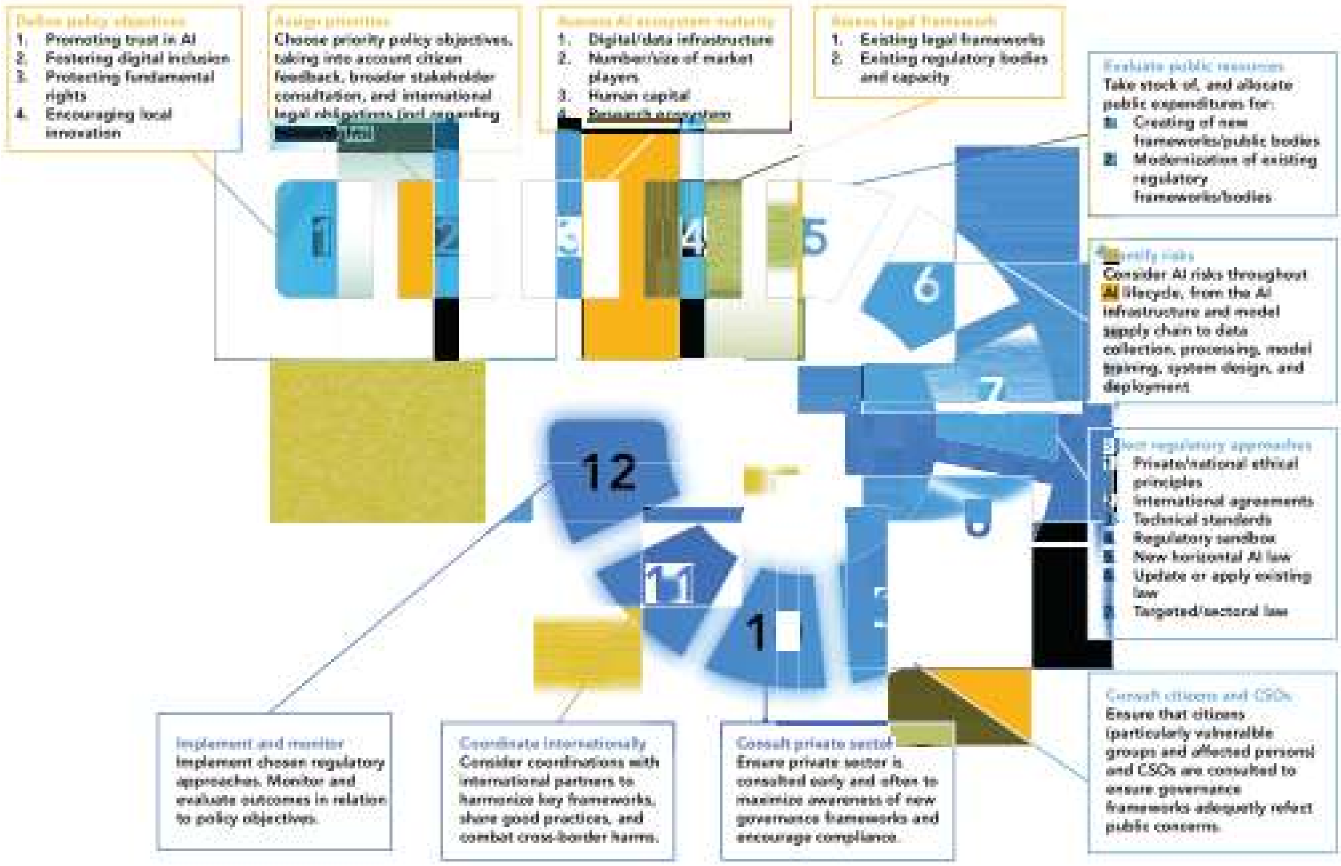
Guidance for Policymakers

ABNASIA.ORG

Trusted AI ecosystems need proportionate, local regulatory frameworks to mitigate risks that arise from AI adoption. As discussed in Section I above, these risks can arise throughout the AI lifecycle – including data privacy and copyright issues that arise during mass data collection, the significant water and energy consumption requirements for model training and inference, and potentially systemic impacts of AI deployment across healthcare, education, social protection, transport, and other critical sectors.

Addressing these risks requires a coordinated, holistic multistakeholder approach to governance. This section is designed to guide policymakers through the stages of developing and implementing effective AI governance interventions. The process allows for continuous evaluation and adjustment to address emerging challenges and opportunities in the AI landscape. AI governance should be a dynamic, agile and adaptive process.

Given the constantly evolving landscape, we do not set out a single prescriptive approach for designing AI governance. The approaches outlined in this section are indicative only and aimed at stimulating policy-level thinking – they are not intended to be a strict rulebook.



Define policy objectives
1. Promoting trust in AI
2. Fostering digital inclusion
3. Protecting fundamental rights
4. Encouraging local innovation

Assign priorities
Choose priority policy objectives, taking into account citizen feedback, broader stakeholder consultation, and international legal obligations (incl regarding

Assess AI ecosystem maturity
1. Digital/data infrastructure
2. Number/size of market players
3. Human capital
4. Research ecosystem

Assess legal framework
1. Existing legal frameworks
2. Existing regulatory bodies and capacity

Evaluate public resources
Take stock of, and allocate public expenditures for:
1. Creating of new frameworks/public bodies
2. Modernization of existing regulatory frameworks/bodies

Identify risks
Consider AI risks throughout AI lifecycle, from the AI infrastructure and model supply chain to data collection, processing, model training, system design, and deployment

Select regulatory approaches
1. Private/national ethical principles
2. International agreements
3. Technical standards
4. Regulatory sandbox
5. New horizontal AI law
6. Update or apply existing law
7. Targeted/sectoral law

Consult citizens and CSOs
Ensure that citizens (particularly vulnerable groups and affected persons) and CSOs are consulted to ensure governance frameworks adequatly refect public concerns.

Consult private sector
Ensure private sector is consulted early and often to maximize awareness of new governance frameworks and encourage compliance.

Coordinate internationally
Consider coordinations with international partners to harmonize key frameworks, share good practices, and combat cross-border harms.

Implement and monitor
Implement chosen regulatory approaches. Monitor and evaluate outcomes in relation to policy objectives.

Figure 7. Process for AI Governance

Source: Authors

ABNASIA.ORG

7.1. Key Considerations

This section is intended to provide high-level guidance for policymakers, outlining 5 key areas for consideration. It is important to note that **there is no single 'one-size-fits-all' regulatory intervention to address AI risks** – regulators should ensure that a combination of the regulatory tools outlined above are adapted to their local country context, consumer needs, and policy priorities.

When selecting the correct regulatory and governance interventions for their consumer's needs and country context, policymakers need to assess at the minimum the following factors:

Table 6. 5 Key considerations before adopting a regulatory approach

Factor	Description
Policy priorities and local context	A country's approach to AI governance will ultimately depend on its policy priorities. These may include promoting trust, protecting fundamental rights of users, increasing digital inclusion, fostering local innovation ecosystems, increasing market competition, or attracting investment.
Maturity of AI ecosystem	Regulatory approaches vary based on the maturity of the local AI ecosystem (including availability of infrastructure and human expertise). In regions with limited AI development, priorities may include promoting good data governance and introducing baseline governance requirements. In more advanced regions, early binding measures to prevent AI risks and harmonization with global standards may be necessary.
Legal framework and regulatory environment	Policymakers need to assess existing legal frameworks to determine which regulatory tools can be implemented immediately and which may require legislative reform. This includes considering AI-specific provisions in data protection, cybercrime, competition, and human rights laws and dispute resolution mechanisms among others.
Public resources and capacity	Policymakers must consider available public resources when designing AI policy interventions. Binding measures require significant investment and skilled staff but create lasting oversight mechanisms. In contrast, self-governance and soft law approaches need less investment but may require some centralized public monitoring.
Stakeholder ecosystem	Effective AI governance requires a comprehensive stakeholder ecosystem, including government bodies, industry participants, academia, civil society, and consumers. Engaging these stakeholders ensures that AI policies are well-rounded, addressing diverse concerns and leveraging collective expertise. Market trust and active participation are crucial for the success of regulatory frameworks.

ABNASIA.ORG

A country's policy priorities are shaped by the unique socio-economic, political, and technological contexts of each country. For instance, a country with a robust tech sector may prioritize fostering innovation and market competition, while a developing country might focus more on digital inclusion and protecting user rights. Moreover, policymakers should ensure that these priorities are grounded in the needs of their consumers. Consumer needs can vary significantly across different demographics and regions, and it is crucial for governance frameworks to reflect these diverse perspectives.

Policy priorities have not been included in the table below as they are specific to each country. The remaining four considerations are included in the analytical matrix to support policymakers' decision-making. This matrix should be read in conjunction with the Dimensions for AI Governance in Section IV and the Tradeoffs for AI Governance in Section III. Note that this matrix is not intended to be exhaustive.

Table 7. Governance Tradeoffs for AI Governance

Regulatory Tool	Maturity of AI ecosystem	Legal framework	Public resources and capacity	Stakeholder ecosystem
Industry self-governance				
Private ethical codes and councils	Suitable for robust industries with established players.	No legal framework required. Public sector can encourage adoption.	Few public sector resources needed for monitoring and encouragement.	Requires trust from industry actors. Limited public input.
Soft Law				
Non-binding international agreements	Suitable for varying sizes and capacities; global harmonizing effect.	No legal framework required at first but can eventually lead to the creation of national laws.	Minimal resources to accede; increases if translated into policy.	Requires cooperation between international and national stakeholders.
National AI principles / ethics frameworks	Suitable for all market sizes; flexible and adaptive.	No explicit legal framework required; new public bodies may be needed.	Varies; can become resource-intensive if new institutions are needed.	Requires broad stakeholder engagement for effective implementation.
Technical standards	Suitable for mature markets with technical capacity.	No legal framework required; can interact with future frameworks.	Varies based on the standard-setting organization and modes of participation. Potentially resource-intensive if setting up new national standard-setting body.	Multi-stakeholder involvement; risk of dominant player influence.

ABNASIA.ORG

Regulatory Tool	Maturity of AI ecosystem	Legal framework	Public resources and capacity	Stakeholder ecosystem
Industry self-governance				
Regulatory sandboxes	Relevant for more developed AI markets with active players.	Requires legal powers to amend regulatory obligations as needed.	Requires substantial resources for establishment, maintenance and monitoring.	High trust needed from market; regulator decisions are discretionary.
Industry self-governance				
New horizontal AI law	Relevant for markets with a clear gap in the regulatory environment. 'Asymmetric' regulatory approach can be adopted, placing greater burden on larger or more systemically risky market players.	Requires new legal framework.	Requires substantial resources to design, implement and oversee new legal and regulatory framework.	Stakeholder buy-in critical; public consultations often required; Coordination between different ministries is essential.
Update or apply existing laws	Most relevant for markets with established actors familiar with existing compliance regimes.	Existing, robust legal frameworks required. Some legislative intervention (at primary or secondary levels) may be needed to modify scope of existing legal regimes or empower existing regulators.	Resources required to modernize regulatory frameworks and increase AI-specific capacity within each regulatory body.	Requires familiarization from regulated entities; co-ordination between different agencies and public trust needed.
Targeted technical or sectoral approaches	Suitable for markets with clear, specific use cases.	New legal frameworks required.	Resource requirements will vary greatly depending on scope and regulatory burdens contemplated by new legal framework.	Requires engagement with specific sectors; high market trust.

Source: authors

ABNASIA.ORG

7.2. Looking to the future

In conclusion, the rapidly evolving landscape of AI presents unique challenges and opportunities for policymakers worldwide. As AI technologies become increasingly integral to various sectors, it is imperative to establish robust regulatory frameworks that can both harness the benefits of AI and mitigate its potential risks. The diverse regulatory tools explored in this paper—including industry self-governance, soft law, national AI principles, technical standards, regulatory sandboxes, and hard law approaches—each offer distinct advantages and limitations. Policymakers must carefully consider the maturity of their AI ecosystem, the existing legal and regulatory environment, public resources, and stakeholder ecosystems when selecting the most appropriate regulatory mechanisms.

Effective AI governance requires a dynamic and flexible approach, allowing for continuous adaptation to new technological developments and societal needs. The proposed Framework provides a structured yet agile process for policymakers to follow, emphasizing the importance of defining clear policy objectives, prioritizing actions based on consumer needs and local contexts, and engaging with a wide range of stakeholders. By fostering collaboration between the public and private sectors, civil society, and international partners, policymakers can ensure that AI governance frameworks are comprehensive, inclusive, and capable of promoting trust, innovation, and ethical standards in AI deployment.

However, the need for governance is underscored by several critical risks associated with AI, though these risks are not exhaustive. AI systems can perpetuate bias and discrimination due to unrepresentative datasets and a lack of transparency in algorithms. The adoption of AI technologies may lead to significant labor market disruption, resulting in job losses and a widening digital divide. Additionally, AI can be misused for spreading misinformation, creating deepfakes, conducting cybercrime, interfering with elections, and facilitating fraud, all of which erode trust in media and news. The environmental impacts of AI are also concerning, as AI systems, particularly those involving large-scale data processing, consume significant amounts of energy, contributing to environmental degradation and increased carbon emissions. Furthermore, cybersecurity vulnerabilities in AI systems and applications are significant due to their complexity and multiple points of vulnerability, highlighting the urgent need for robust regulatory frameworks.

Ultimately, the goal of AI governance should be to create a balanced and forward-looking framework that protects fundamental rights, promotes digital inclusion, and drives sustainable innovation. As AI continues to transform our world, it is crucial that regulatory approaches are not only effective but also equitable and responsive to the diverse needs of all stakeholders. Specific policy recommendations should be carefully considered and tailored to the context of each country. Upcoming papers on prerequisites for AI and AI toolkits for creating strategies will further support policymakers in their efforts to build effective AI governance frameworks. By leveraging the insights and tools discussed in this paper as a starting point for broader policy discussion and stakeholder consultation, policymakers can navigate the complexities of AI governance and build a foundation for a safer, more inclusive, and innovative future.

ABNASIA.ORG

GLOSSARY

AI (Artificial Intelligence): A branch of computer science focused on creating systems capable of performing tasks that typically require human intelligence, such as decision-making, visual perception, speech recognition, and language translation.

AI Ethics: The study of the ethical and moral implications of AI, focusing on ensuring that AI technologies are developed and used in ways that are fair, transparent, and accountable.

AI Governance: The framework of laws, rules, practices, and processes used to ensure AI technologies are developed and used responsibly.

AI Regulation: Binding legal and regulatory frameworks enacted to influence AI development and deployment.

AI Systems: A machine-based system that infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

Algorithmic Bias: The systematic and repeatable errors in a computer system that create unfair outcomes, such as privileging one arbitrary group of users over others.

Compute Capacity: The ability to store, process, and transfer data at scale, which is crucial for training and deploying AI models and applications.

Cybersecurity: The practice of protecting systems, networks, and programs from digital attacks.

Data Governance: The management of data availability, usability, integrity, and security in an enterprise or organization. This includes data privacy and cybersecurity laws and regulations.

Data Privacy: The right of individuals to control how their personal information is collected and used.

Deep Learning: A subset of ML involving neural networks with many layers (hence 'deep') that can learn from large amounts of data.

Deepfakes: Synthetic media where a person in an existing image or video is replaced with someone else's likeness using AI.

Digital Inclusion: Efforts to ensure that all individuals and communities, including the most disadvantaged, have access to and can use information and communication technologies.

Digital Literacy: The ability to use information and communication technologies to find, evaluate, create, and communicate information, requiring both cognitive and technical skills.

Digital Public Infrastructure (DPI): Digital platforms for identity, payments, and data sharing that are foundational for accessing services and boosting digital inclusion.

Environmental Impact of AI: The effects that the development, training, and deployment of AI systems have on the environment, including energy consumption and carbon emissions.

Generative AI: A type of AI that can generate new content, such as text, audio, images, and video, based on the data it has been trained on. Examples include OpenAI's GPT series and Meta's Llama.

Generative AI Models: AI systems that can create new content. Examples include OpenAI's GPT series, Anthropic's Claude, Google DeepMind's Gemini, and Meta's Llama.

High-Quality Data: Data that is accurate, complete, reliable, relevant, and timely, essential for effective AI training and deployment.

Human Capital: The skills, knowledge, and experience possessed by an individual or population, viewed in terms of their value or cost to an organization or country.

Intellectual Property (IP): Legal rights that result from intellectual activity in the industrial, scientific, literary, and artistic fields. In AI, this includes patents, copyrights, and trademarks related to AI technologies.

ABNASIA.ORG

Machine Learning (ML): A subset of AI that involves the use of algorithms and statistical models to enable computers to improve their performance on a specific task with experience.

Narrow AI: Also known as traditional AI, designed to perform a specific task such as facial recognition or fraud detection. It operates based on explicit programming and rules.

OECD AI Principles: Guidelines set by the Organisation for Economic Co-operation and Development to promote innovative and trustworthy AI that respects human rights and democratic values.

Public-Private Partnership (PPP): A cooperative arrangement between one or more public and private sectors, typically of a long-term nature.

Regulatory Sandbox: A framework set up by a regulator that allows small-scale, live testing of innovations under a regulator's oversight.

Regulatory Trade-offs: The balancing of benefits and risks when creating regulations, ensuring innovation is not stifled while protecting public interest.

Sustainable AI: AI that is developed and deployed in a manner that is environmentally, economically, and socially sustainable.

ABNASIA.ORG

ANNEX: SAMPLE COUNTRY APPROACHES TO AI GOVERNANCE

This Annex is intended to provide a high-level snapshot of how certain countries have built AI governance frameworks by combining the regulatory tools outlined in Section 4.

Brazil

Soft Law / Regulatory Sandbox

In October 2023, Brazil's data protection authority (the ANPD) launched a regulatory sandbox pilot program for AI and data protection – this included a public consultation process with the public and private sectors.[218] Although the impact of the Brazilian regulatory sandbox on the local innovation ecosystem and regulatory compliance is not yet clear due to its early stage, it is notable for a broad multi-stakeholder approach, seeking to coordinate action between regulators, regulated entities, technology companies, academics and civil society organizations.[219]

Brazil has endorsed both the OECD and G20 AI Principles and has referenced the OECD Principles as guidance for developing its own national AI strategy. Brazil has also joined the Global Partnership on AI.

Brazil has also endorsed the UNESCO Recommendation on the Ethics of AI. Brazil was one of the first countries to complete the UNESCO Readiness Assessment Methodology.[220] Brazil is also a signatory to the 2023 Santiago Declaration to Promote Ethical Artificial Intelligence,[221] which reflects the UNESCO Recommendation.

As a member of the Latin American Centre for Development Administration, Brazil approved the Iber American Charter on Artificial Intelligence in Civil Service in November 2023.[222] The Charter is a non-binding roadmap of best practices for states to guide the implementation of AI in public administration, and emphasizes de-biasing AI systems, improving transparency, protection fundamental rights, and improving public trust in AI. The charter suggests creation of domestic public registry of algorithms used in the public sector, and the establishment of public oversight, audit, and risk assessment mechanisms.

Hard Law

Brazil's Bill 2.338/2023 proposes a risk and rights-based approach to AI governance. It proposes classifying AI systems into three levels of risk: (i) excessive risk, in which the use is prohibited; (ii) high risk; and (iii) non-high risk. AI systems should pass a preliminary self-assessment analysis conducted by the AI provider to classify its risk level. Every AI system must implement a governance structure involving transparency, data governance and security measures.[223] In addition, high-risk AI systems must include technical documentation, log registers, reliability tests, technical explainability measures and measures to mitigate discriminatory biases.[224]

The Bill proposes individual rights, such as the right to explanation about decisions, non-discrimination and correction of discriminatory biases, and the right to privacy and protection

[218] ANPD's Call for Contributions to the regulatory sandbox for artificial intelligence and data protection in Brazil is now open, GOV.BR (Oct. 3, 2023), https://www.gov.br/anpd/pt-br/assuntos/noticias/anpds-call-for-contributions-to-the-regulatory-sandbox-for-artificial-intelligence-and-data-protection-in-brazil-is-now-open.

[219] Brazil: ANPD opens AI regulation sandbox for public consultation, ONETRUST DATAGUIDANCE (Oct. 4, 2023), https://www.dataguidance.com/news/brazil-anpd-opens-ai-regulation-sandbox-public.

[220] https://www.unesco.org/ethics-ai/en/brazil

[221] https://minciencia.gob.cl/uploads/filer_public/40/2a/402a35a0-1222-4dab-b090-5c81bbf34237/declaracion_de_santiago.pdf

[222] https://clad.org/wp-content/uploads/2024/03/CIIA-EN-03-2024.pdf

[223] Id

[224] https://accesspartnership.com/access-alert-brazils-new-ai-bill-a-comprehensive-framework-for-ethical-and-responsible-use-of-ai-systems/; https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-brazil

ABNASIA.ORG

of personal data.[225] The Bill also includes rules for civil liability, codes of best practice, notification of AI incidents, copyright exceptions for data mining processing, and fostering of regulatory sandboxes. Furthermore, it proposes the creation of an open public database of high-risk AI systems that contains public documentation of algorithmic impact assessments. The Executive Branch is tasked to designate a supervisory authority to regulate and enforce legislation regarding Brazil's National AI Strategy (EBIA).

Brazil's national data protection authority, the ANDP, has also adopted a Resolution CD/ANPD No 10 on strengthening data protection and oversight of AI applications.[226] As a member of the Iber-American Network for the Protection of Personal Data, the ANDP has also endorsed the region-wide General Recommendations for the Processing of Personal Data in Artificial Intelligence.[227]

The ANDP has been active in taking regulatory action against AI developers. For example, in July 2024 the ANDP took regulatory action to suspend Meta's latest privacy policy, preventing it from using Brazilians' Instagram and Facebook posts to train its AI models.[228]

United States

Soft Law / Regulatory Sandbox

The National Institute of Standards and Technology (NIST) AI Risk Management Framework (RMF) provides guidance for risk mitigation across the value chain.[229] NIST convenes multistakeholder experts to develop guidance for generative AI and on safety concerns such as synthetic content, capability evaluations, red-teaming of AI systems, biosecurity and cybersecurity risks for foundation models.[230]

NIST also houses the US AI Safety Institute, a consortium of over 200 leading AI stakeholders including AI creators and users, academics, government and industry researchers, and civil society organizations, which aims to advance the development of safe, trustworthy AI. The AI Safety Institute contributes to the priority actions outlined in the administration's Executive Order, including developing guidelines for red-teaming, capability evaluations, risk management, safety and security, and watermarking synthetic content.[231]

In 2022 the Office of Science and Technology Policy (OSTP) of the White House produced a 'Blueprint for an AI Bill of Rights' suggesting fundamental principles to guide and govern the efficient development and implementation of AI systems. These include the following:

1. Safe and effective systems: Users should be protected from unsafe or ineffective systems.

2. Algorithmic discrimination protections: users should not be exposed to discrimination by algorithms; automated decision-making systems should be used and designed equitably.

3. Data privacy: users should be protected from abusive data practices via built-in protections and have agency over how their data is used.

4. Notice and explanation: users must be informed that an automated system is being used and understand how and why it contributes to outcomes that impact them.

5. Alternative options: users should have the right to opt out, where appropriate, and have access to a person who can quickly consider and remedy their problems.[232]

[225] https://accesspartnership.com/access-alert-brazils-new-ai-bill-a-comprehensive-framework-for-ethical-and-responsible-use-of-ai-systems/

[226] https://www.gov.br/anpd/pt-br/documentos-e-publicacoes/documentos-depublicacoes/nota-tecnica-no-19-2023-fis-cgf-anpd.pdf

[227] https://www.redipd.org/sites/default/files/2020-02/guide-generalrecommendations-processing-personal-data-ai.pdf

[228] https://www.bbc.com/news/articles/c7291l3nvwvo

[229] https://www.nist.gov/itl/ai-risk-management-framework

[230] https://www.nist.gov/artificial-intelligence/artificial-intelligence-safety-institute/aisic-working-groups

[231] https://www.commerce.gov/news/press-releases/2024/02/biden-harris-administration-announces-first-ever-consortium-dedicated

[232] https://www.whitehouse.gov/ostp/ai-bill-of-rights/

ABNASIA.ORG

Hard Law

The United States does not have any comprehensive federal legislation on AI, or a single national AI governance strategy; its emerging AI governance regime is composed from various pieces of state-level legislation, along with national principles, guidance, and policies.

In October 2023 the U.S. President signed an Executive Order directing federal agencies to update their mandates for ensuring safe, secure and trustworthy AI - on topics ranging from biosecurity and cybersecurity to discrimination and international development.[233] Accordingly, agencies have been progressing in 2024 to meet their objectives.[234] For example, the National Telecommunications and Information Administration (NTIA), housed within the Department of Commerce, has published the Artificial Intelligence Accountability Policy Report, suggesting independent audits and certifications, funding for red-teaming and evaluations, applying liability laws, transparency disclosures and reporting for incidents and information about models and their training, and consequences for imposing unacceptable risks or making unfounded claims.[235] This report and NTIA's forthcoming guidance on open source AI risk mitigation were informed by public requests for comment or information.

Several US sectoral regulators have begun clarifying the scope of their regulatory authority over AI. The United States the Federal Trade Commission (FTC) has clarified that they will exercise powers against 'unfair and deceptive practices,' fraud, scams,[236] deception, including impersonations generated by AI.[237] It has issued a resolution for civil investigative demands into AI products.[238] The U.S. Securities and Exchange Commission indicated they will address AI and predictive data analytics in finance and investing.[239] The Consumer Financial Protection Bureau (CFPB) is clarifying how existing federal anti-discrimination law applies to algorithmic systems used for lending decisions[240] and has published a report on the risks and use of Chatbots in Consumer Finance.[241] The Equal Employment Opportunity Commission (EEOC) has provided guidance for anti-discrimination laws related to algorithm-based hiring.[242] In addition, the proposed Algorithmic Accountability Act of 2022 would direct the FTC to develop impact assessments of automated ML decision-making processes.[243]

The US has also sought to govern AI through the imposition of a range of more targeted governance measures. For example, the US Department of Commerce, Bureau of Industry Security (BIS) has introduced a range of export control measures aimed at restricting the export of advanced semiconductors and other related equipment to China and other countries.[244]

[233] 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/

[234] https://www.whitehouse.gov/briefing-room/statements-releases/2024/03/28/fact-sheet-vice-president-harris-announces-omb-policy-to-advance-governance-innovation-and-risk-management-in-federal-agencies-use-of-artificial-intelligence/

[235] https://www.ntia.gov/sites/default/files/publications/ntia_ai_report_final-3-27-24.pdf

[236] https://www.ftc.gov/business-guidance/blog/2023/02/keep-your-ai-claims-check

[237] https://www.ftc.gov/news-events/news/press-releases/2024/02/ftc-proposes-new-protections-combat-ai-impersonation-individuals

[238] https://www.ftc.gov/news-events/news/press-releases/2023/11/ftc-authorizes-compulsory-process-ai-related-products-services

[239] https://www.sec.gov/news/testimony/gensler-testimony-house-financial-services-041823

[240] https://www.consumerfinance.gov/about-us/newsroom/cfpb-acts-to-protect-the-public-from-black-box-credit-models-using-complex-algorithms/

[241] https://www.consumerfinance.gov/data-research/research-reports/chatbots-in-consumer-finance/chatbots-in-consumer-finance/

[242] https://www.eeoc.gov/laws/guidance/select-issues-assessing-adverse-impact-software-algorithms-and-artificial

[243] https://www.congress.gov/bill/117th-congress/house-bill/6580/text

[244] https://www.nortonrosefulbright.com/en/knowledge/publications/5a936192/us-expands-export-restrictions-on-advanced-semiconductors

ABNASIA.ORG

China

Soft Law / Regulatory Sandbox

National standard-setting bodies have begun to play key role in formulating standards to facilitate the implementation of the legal frameworks outlined above. For example, the National Information Security Standardisation Technical Committee of China ('TC260') released TC260-003 Basic security requirements for generative artificial intelligence service,[245] which provides companies with practical guidance on complying with the 2023 Generative AI Measures' requirements regarding training data security, model security, internal governance measures, and the conducting of security assessments.

China has also recently announced that it will launch a 'Global AI Governance Initiative' in a keynote speech at the Opening Ceremony of the third Belt and Road Forum on International Cooperation. The official government statement notes that it supports discussions with the UN framework to establish 'an international institution to govern AI, and to coordinate efforts to address major issues concerning international AI development, security and governance.' While the details of the initiative are not clear, the press release issued by the government provides that the focus will be on China's proposals on AI governance regarding the development, security and governance of AI.

Hard Law

China was one of the first jurisdictions to introduce any form of legislation specifically governing AI – it currently has separate laws regulating recommendation algorithms, 'deep synthesis' technologies (a subset of generative AI technologies that includes deepfakes and digital simulation models), and generative AI services.[246] Chinese policymakers have also indicated that they will seek to formulate a general, horizontal AI law in the coming years.[247]

Law firm Bird & Bird has identified three main pillars of China's overall AI governance regime:[248]

1. Content moderation: The first pillar of China's AI regulatory regime concerns the governance and management of online content. With respect to AI-generated content (such as the output text of an LLM), regulators will prioritize traceability and authenticity of the content to restrict circulation of information that would violate well-established information services regulations.
2. Data protection: Data protection is governed by the 2021 Personal Information Protection Law, which aims to ensure that personal data processing does not harm users or otherwise undermine public order. The PIPL enshrines key principles including lawfulness of processing, transparency, sincerity, and accountability.
3. Algorithmic governance: Security assessments play a key role in Chinese AI regulation; administered by the CAC, these assessments involve complex filing procedures and require listing of in-scope algorithms on an online registry (particularly for services with 'public opinion attributes or social mobilization capabilities.') Chinese AI regulation also seeks to ensure that AI services reflect 'public order and morality', e.g. the CAC prohibits the use of AI to generate any discriminatory content or decision based on race, ethnicity, beliefs, nationality, region, gender, age, occupation, and health.[249] In addition, AI services that generate human-like content (whether textual, visual, or auditory) must present clear, specific and actionable annotation rules and make clear that content has been generated with the use of AI.[250] The Generative AI Measures directly regulate model training practices by requiring service providers to use data and models from

[245] https://www.tc260.org.cn/upload/2024-03-01/1709282398070082466.pdf

[246] https://www.lw.com/admin/upload/SiteAttachments/Chinas-New-AI-Regulations.pdf

[247] https://www.gov.cn/zhengce/content/202306/content_6884925.htm

[248] https://www.twobirds.com/en/insights/2024/china/ai-governance-in-china-strategies-initiatives-and-key-considerations#:~:text=China's%20AI%20governance%20framework%20attempts,to%20make%20decisions%20about%20individuals

[249] Article 4(2), 2023 Generative AI Measures.

[250] Articles 8 and 12, 2023 Generative AI Measures.

ABNASIA.ORG

legitimate sources, respect intellectual property rights and personal information, and strive to improve the quality, authenticity, accuracy, objectivity and diversity of the training data they utilize.[251]

China's AI governance measures are formulated to promote China's specific policy interests and national priorities – for example, the Generative AI Measures require generative AI service providers to uphold 'socialist core values' and prohibits the generation of certain types of content, such as content that incites 'subversion of the state power or the overthrow of the socialist system, endangers national security and interests, damages the national image, incites splitting the country, undermines national unity and social stability, advocates terrorism, extremism, ethnic hatred and discrimination, violence, pornography, and false and harmful information'.[252]

In addition to the use-case focused regulatory frameworks outlined above, regional regulations have been used in China to promote local AI development and create local-level experiments to attract AI investment. Enacted in 2022, the Shanghai Regulations on Promoting the Development of AI Industry 2022 and the Shenzhen Special Economic Zone Regulations on AI Industry Promotion 2022 both call for the creation of AI Ethics Committees to oversee AI development, conduct audits and assessments, and promote industrial parks where input and training data may be traded easily and lawfully.[253]

United Kingdom

Soft Law / Regulatory Sandbox

The United Kingdom government introduced its cross-sector plan for AI regulation on July 18, 2022, which features a 'pro-innovation' framework. These non-statutory principles apply broadly and are supplemented by 'context-specific' regulatory guidance and voluntary standards developed by UK regulators. The UK is moving towards a light-touch, risk-based, context-specific approach focused on proportionality, with practical requirements determined by the industry and dependent on the AI system's deployment context.[254] The Alan Turing Institute, as the national institute for data science and AI, plays a pivotal role in research and ethics in AI. In February 2024, the UK Department for Science, Innovation and Technology updated its 'A pro-innovation approach to AI regulation' after a public consultation.[255]

Outside of the 'pro-innovation' regulatory framework, the UK government has also adopted the following policy tools:

- The government has announced a Foundation Model Taskforce which has been allocated £100 million in funding and will focus on accelerating the UK's capability to develop 'safe and reliable' foundation models.
- The UK AI Safety Institute (discussed at box 21 above).
- The UK has also developed an AI Standards Hub to share knowledge, capacity, and research on AI standards.[256]
- The UK hosted the highly publicized AI Safety Summit on 1 - 2 November 2023.

[251] Article 7, 2023 Generative AI Measures.

[252] Article 4(1), 2023 Generative AI Measures.

[253] https://www.twobirds.com/en/insights/2024/china/ai-governance-in-china-strategies-initiatives-and-key-considerations#:~:text=China's%20AI%20governance%20framework%20attempts,to%20make%20decisions%20about%20individuals

[254] UK Government (2022a)

[255] https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response#a-regulatory-framework-to-keep-pace-with-a-rapidly-advancing-technology

[256] https://aistandardshub.org/

ABNASIA.ORG

The Summit was attended by a number of countries, as well as companies and a small selection of civil society organizations. The Bletchley Declaration by countries in attendance at the AI Safety Summit refers to ensuring wider international cooperation on AI and sustaining an inclusive global dialogue that engages existing international fora and other relevant initiatives and contributes in an open manner to broader international discussions.

Hard Law

The UK has stated that it will harness its existing regulators through cross-sector legislation rather than setting up a new regulator. At the same time, the UK has recognized that a patchwork approach to regulation with little central coordination or oversight may in fact create barriers to innovation due to a lack of coherence and clarity in regulatory obligations – as such, the UK has committed to create a set of centralized mechanisms to ensure the sectoral approach to AI regulation can be monitored and adapted, as well as facilitate a single point of collaboration for all interested parties (international partners, industry, civil society, academia and the public) (see figure 7 above).

One coordination mechanism is the Digital Regulation Cooperation Forum (DRCF) Multi-Agency Advisory Service, a pilot scheme that will see a number of regulators develop a multi-agency advice service providing tailored support to businesses using AI and digital innovations so they can meet requirements across various sectors.[257]

Sectoral regulators in the UK have already begun to clarify the scope of their mandates as they relate to AI:

- The UK Competition and Markets Authority published an Initial Report on AI Foundation Motels in September 2023, which was supplemented by an 'Update Paper' published in April 2024. These reports examine the CMA's understanding of AI risks, how the CMA's competition and consumer remit applies to those AI risks, forthcoming changes to the CMA's powers and the CMA's AI capabilities.[258] These reports were published after broad stakeholder consultation with consumer groups, civil society, leading AI developers and deployers, academics, and other regulators.[259]
- In April 2024 the UK data protection regulator, the Information Commissioner's Office (ICO), published its strategic approach to regulating AI, which sets out how the ICO is driving forward the principles set out in the UK government's AI regulation white paper.[260] Although the UK government has not appointed a separate AI regulator, the ICO notes that many of the principles identified in the UK's AI regulation white paper align with established data protection principles, meaning that the ICO may eventually become a de facto AI regulator.

At the same time, discussions on introducing a formal regulatory framework for AI in the UK have gained momentum. In November 2023, the Artificial Intelligence (Regulation) Bill was introduced in the House of Lords.[261] The Bill proposes a central AI authority which will ensure alignment of approach by different regulators, as well as ensuring that relevant regulators take account of AI, monitoring the effectiveness of the UK AI framework and collaborate with regulators to construct regulatory sandboxes for AI. The Bill also requires AI developers to comply with requirements regarding transparency, IP rights, and labelling of AI outputs. It also establishes the role of 'AI responsible officers' for businesses that develop, deploy, or use AI.

[257] https://www.drcf.org.uk/home

[258] https://www.gov.uk/government/publications/ai-foundation-models-initial-report; https://www.gov.uk/government/publications/cma-ai-strategic-update/cma-ai-strategic-update

[259] id

[260] https://www.skadden.com/-/media/files/publications/2024/05/the-uk-ico-publishes-its-strategy-on-ai-governance/regulating-ai-the-icos-strategic-approach.pdf?rev=7752a638c485400fb9e1e84dbe077ab6&hash=16CEEC36DAF5CD2A68D4F7A14F30A401

ABNASIA.ORG

India

Soft Law / Regulatory Sandbox

India's National Strategy for AI (published in June 2018) identified a lack of formal regulation around data as a key barrier for large scale adoption of AI.[262]

The Principles for Responsible AI (adopted in February 2021) serve as India's roadmap for creating a responsible AI ecosystem across sectors. It identifies the following relevant principles:

1. The principle of safety and reliability
2. The principle of equality
3. The principle of inclusivity and non-discrimination
4. The principle of privacy and security
5. The principle of transparency
6. The principle of accountability
7. The principle of protection and reinforcement of positive human values

The Operationalizing Principles for Responsible AI (August 2021) identifies actions that need to be taken by both the government and the private sector, in partnership with research institutes, to cover regulatory and policy interventions, capacity building, incentivizing ethics by design, and creating frameworks for compliance with relevant AI standards.

Indian sectoral regulators have issued guidance on the regulation of AI.[263]

1. In the finance sector, the Securities and Exchange Board of India issued a circular in January 2019 on reporting requirements for AI and machine learning applications and systems offered and used.
2. In the health sector, the strategy for National Digital Health Mission identifies the need for the creation of guidance and standards to ensure the reliability of AI systems in health.

The Indian Ministry of Electronics & Information Technology has established four committees on AI, which have published several reports on security, safety, legal and ethical issues relating to AI.[264]

India is a member of the Global Partnership on Artificial Intelligence (GPAI). The 2023 GPAI Summit was recently held in New Delhi, where GPAI experts presented their work on responsible AI, data governance, and the future of work, innovation, and commercialization.[265]

The Bureau of Indian Standards, the national standards body of India, has established a committee on AI that is proposing draft Indian standards for AI.[266]

India is a party to the OECD's AI principles and has adopted UNESCO's Recommendation on the Ethics of AI.[267]

Hard Law

India does not currently have any horizontal AI law.

However, the proposed Digital India Act, replacing the IT Act of 2000, may regulate AI systems. Although the Act is primarily intended to be a form of internet platform regulation, the proposed Act intends to regulate high-risk systems through 'legal, institutional quality testing framework to examine regulatory models, algorithmic accountability, zero-day threat & vulnerability assessment, examine AI based ad-targeting, content moderation etc.'[268]

India also recently concluded its first data protection law, the Digital Personal Data Protection Act 2023 – however, as of the time of publication, the law has yet to come into force.[269]

[261] https://www.engage.hoganlovells.com/knowledgeservices/news/new-uk-regulation-bill-potential-step-forward-to-the-statutory-regulation-of-ai-systems-in-the-uk#:~:text=The%20primary%20purpose%20of%20the,the%20regulatory%20approach%20to%20AI

[262] https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf

[263] https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-india

[264] https://www.meity.gov.in/artificial-intelligence-committees-reports

[265] https://www.morganlewis.com/blogs/sourcingatmorganlewis/2024/01/ai-regulation-in-india-current-state-and-future-perspectives

[266] https://www.services.bis.gov.in/php/BIS_2.0/dgdashboard/Published_Standards_new/standards?commtid=Mzg2&commttname=TEIURCAzMA%3D%3D&aspect=&doe=&from=2022-07-21&to=2023-07-21

[267] https://iapp.org/media/pdf/resource_center/global_ai_law_policy_tracker.pdf

[268] https://www.meity.gov.in/writereaddata/files/DIA_Presentation%2009.03.2023%20Final.pdf

[269] https://www.dlapiperdataprotection.com/index.html?t=law&c=IN#:~:text=Under%20the%20DPDP%20Act%2C%20Data,necessary%20for%20the%20specified%20purpose

ABNASIA.ORG

Nigeria

Soft Law / Regulatory Sandbox

In August 2024 the National Information Technology Development Agency's (NITDA) National Center for Artificial Intelligence and Robotics (NCAIR) published a draft National Artificial Intelligence Strategy (NAIS).[270] Two pillars of the strategy that touch on governance issues are highlighted below:

Pillar 4: Ensuring Responsible and Ethical AI Development.

Under this pillar, Nigeria aims to:

1. Create a high-level AI ethics expert group / national ethics commission comprised of stakeholders from academia, industry, government, and civil society, to develop and implement ethical AI principles.
2. Develop national AI ethical principles that align with critical Nigerian values.
3. Develop a comprehensive AI ethics assessment framework
4. Implement legislative reforms to address emerging legal and ethical challenges

Pillar 5: Developing a Robust AI Governance Framework

Under this pillar, Nigeria aims to:

1. Develop national AI principles to guide development, deployment and use of AI
2. Establish an AI governance regulatory body to oversee implementation of the national AI principles, ensure compliance with ethical standards, and mediate potential disputes.
3. Develop a national AI policy framework that defines governance guidelines and principles for AI systems
4. Develop a national AI risk management framework

The NAIS also identifies the US NIST Framework for AI Risk Management as a valuable tool for guiding the design and deployment of AI systems.

Hard Law

Nigeria currently has not proposed any AI legislation. However, law firm White & Case has identified several existing laws that affect the development or use of AI in Nigeria, including:[271]

1. The Cybercrimes (Prohibition, Prevention, etc.) Act, 2015
2. The Nigeria Data Protection Act, 2023
3. The Security and Exchange Commission (SEC) Rules on Robo-Advisory Services
4. The Federal Competition and Consumer Protection Act, 2018
5. The Copyright Act, 2022
6. The Nigerian Communication Commission Act, 2003

In addition, Pillar 4 of the draft National AI Strategy notes that legal reforms may be needed to address particular legal or ethical concerns arising from AI, including 'protecting workers' rights through retraining programs, tailored unemployment benefits, and policies encouraging job sharing and reduced work hours. Additionally, bridging the digital divide requires legislation promoting digital literacy and equitable access to technology'.[272]

[270] https://ncair.nitda.gov.ng/wp-content/uploads/2024/08/National-AI-Strategy_01082024-copy.pdf

[271] https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-nigeria#:~:text=There%20is%20currently%20no%20specific,Artificial%20Intelligence%20Policy%20(NAIP)

[272] https://ncair.nitda.gov.ng/wp-content/uploads/2024/08/National-AI-Strategy_01082024-copy.pdf

ABNASIA.ORG

Singapore

Soft Law / Regulatory Sandbox

Singapore has developed a range of voluntary governance frameworks for ethical AI deployment. This includes the Model AI Governance Framework (2019, updated in 2020) which provides detailed guidance to private sector organizations to address key ethical and governance issues when deploying AI solutions.[273]

In response to the growing adoption of generative AI, the AI Verify Foundation and IMDA published a Model AI Governance Framework for Generative AI in May 2024, emphasizing the need for building a trusted ecosystem for AI, highlighting unique risks that arise from generative AI (e.g. hallucination, copyright infringement, value alignment) and emphasizing the need for balancing user protection and innovation.[274] The nine dimensions of the framework include:

1. Accountability – allocation of responsibility to players along the AI development chain
2. Data – ensuring quality of data fed to AI models through the use of trusted data sources
3. Trusted development and deployment – encouraging transparency and disclosure to enhance broader awareness and safety
4. Incident reporting – establishing incident-management structures and processes for timely notification and remediation
5. Testing and assurance – adopting third-party testing against common AI testing standards to demonstrate trust to end-users
6. Security – addressing risks of new threat vectors being injected through AI models
7. Content provenance – developing technologies to enhance transparency about where and how content is generated
8. Safety and alignment research & development – accelerating investment in research & development to improve model alignment with human intention and values
9. AI for public good – harnessing AI to benefit the public by democratizing access, improving public sector adoption, upskilling workers and developing AI systems sustainably

Singapore's AI Governance testing framework and toolkit, 'AI Verify,' launched as a pilot in May 2022, validates the performance of AI systems against a set of internationally recognized principles and frameworks through standardized tests. It provides a testing report that serves to inform, users, developers, and researchers. The Future of Privacy Forum notes that, 'rather than defining ethical standards, AI Verify provides verifiability by allowing AI system developers and owners to demonstrate their claims about the performance of their AI systems.'[275]

Singapore has previously experimented with sector-specific regulatory sandboxes for AI. In 2017, a 5-year regulatory sandbox was created to facilitate the safe development and integration of autonomous vehicles.[276]

Hard Law

Singapore does not have a horizontal AI law at present. However, it has several sectoral laws applicable to AI, including:[277]

1. The Road Traffic Act 1961, which was amended in 2017 to allow for the testing and use of autonomous motor vehicles.
2. The Health Products Act 2007, which requires medical devices that incorporate AI technology to be registered before they are used.

Sectoral regulators have begun issuing non-binding guidance on the use of AI in specific industries.[278] The Monetary Authority of Singapore issued the Principles to Promote

[273] https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf

[274] https://aiverifyfoundation.sg/wp-content/uploads/2024/05/Model-AI-Governance-Framework-for-Generative-AI-May-2024-1-1.pdf

[275] Josh Lee Kok Thong, AI Verify: Singapore's AI Governance Testing Initiative Explained, FUTURE OF PRIVACY FORUM (June 6, 2023), https://fpf.org/blog/ai-verify-singapores-ai-governance-testing-initiative-explained/

[276] https://www.ippapublicpolicy.org/file/paper/5cea683b9a45b.pdf

[277] https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-singapore

[278] https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-singapore

ABNASIA.ORG

Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector8 in 2018 (updated in 2019) to provide a set of foundational principles for firms to consider when using AI in decision-making in the provision of financial products and services. The Ministry of Health, Health Sciences Authority and Integrated Health Information Systems jointly issued the Artificial Intelligence in Healthcare Guidelines in 2021 to improve the understanding, codify good practice and support the safe growth of AI in healthcare

One important law is the Personal Data Protection Act. In March 2024, Singapore's Personal Data Protection Commission (PDPC) issued Advisory Guidelines on the Use of Personal Data in AI Recommendation and Decision Systems,[279] providing organizations with clarity on the use of personal data at three stages of AI system implementation: (a) development, testing and monitoring, (b) deployment, and (c) procurement.[280]

Rwanda

Soft Law / Regulatory Sandbox

In May 2023 Rwanda adopted its first National Artificial Intelligence Policy. The drafting of the policy was led by Rwanda's Ministry of ICT and Innovation (MINICT) and Rwanda Utilities Regulatory Authority (RURA) and supported by GIZ FAIR Forward and The Future Society as an implementation partner.

The policy identifies six priority policy areas (Figure 8).

Figure 8. Rwanda National AI Policy

Source: https://www.minict.gov.rw/index.php?eID=dumpFile&t=f&f=67550&token=6195a53203e197efd47592f40f4ad24579640e

[279] https://www.pdpc.gov.sg/guidelines-and-consultation/2024/02/advisory-guidelines-on-use-of-personal-data-in-ai-recommendation-and-decision-systems

[280] https://www.dataprotectionreport.com/2024/03/singapore-releases-new-guidelines-on-the-use-of-personal-data-in-ai-systems/

ABNASIA.ORG

Key actions arising from the report include the following:

1. Strengthen AI policy and regulation, build capacity of regulatory authorities, and ensure public trust in AI
2. Operationalize and share Rwanda's 'Guidelines on the Ethical Development and Implementation of AI', led by RURA.
3. Actively contribute to shaping responsible AI principles & practices in international platforms

Hard Law

Rwanda does not have any binding AI laws at present. However, there are several existing legal frameworks that could apply to AI systems. These include the following:

1. The Law n°058/2021 of 13/10/2021 relating to the protection of personal data and privacy
2. The Law n° 24/2016 of 18/06/2016 governing Information and Communication Technologies in Rwanda
3. The Law n° 60/2018 of 22/8/2018 on prevention and punishment of cyber-crimes

UAE[281]

Soft Law / Regulatory Sandbox

In 2017, the Minister of State for Artificial Intelligence Office adopted a National Strategy for Artificial Intelligence 2031.

The UAE has adopted a set of non-binding AI Ethics Principles, aiming to ensure responsible and ethical use of AI. Key principles include:[282]

1. Transparency: Ensuring that AI systems and their decision-making processes are understandable and accessible to users.
2. Accountability: Establishing clear lines of responsibility for the development and use of AI systems.
3. Fairness: Mitigating bias in AI systems to ensure equitable treatment of all individuals and groups.

Digital Dubai, a government platform established in 2021, has also released an Ethical AI Toolkit for businesses to use for practical guidance, including a self-assessment tool.[283]

A range of non-binding national guidelines have been issued in relation to AI, including:[284]

1. The Deepfake Guide (2021) – sets out information on deepfakes, and provides advice on measures to protect against deepfakes and guidance on how to report deepfakes to the appropriate authorities
2. The AI Ethics Guide (2022) – sets out non-mandatory guidelines with respect to the ethical design and deployment of AI systems in both the public and private sectors
3. The AI Adoption Guideline in Government Services (2023) – aims to create awareness, accelerate AI impact and to provide a continuously updated repository of clear use cases with respect to the deployment of AI in government services
4. The Responsible Metaverse Self-Governance Framework (2023) – a whitepaper which seeks to establish common minimum self-regulatory principles with respect to responsible use in the metaverse
5. The Guidelines for Financial Institutions adopting Enabling Technologies – issued by the financial services regulators in the Financial Free Zones and in Mainland UAE; suggests governance frameworks for a variety of emerging technologies, including 'big data analytics and artificial intelligence'.

[281] Note that the UAE comprises multiple legal jurisdictions – for the purposes of this Annex, these are categorized as follows: (a) the Financial Free Zones (Dubai International Financial Centre (DIFC) and Abu Dhabi Global Market (ADGM), and (b) the Mainland UAE (the remainder of the UAE outside the FFZ.

[282] https://insight.thomsonreuters.com/mena/legal/posts/how-is-ai-regulated-in-the-uae-what-lawyers-need-to-know

[283] https://www.digitaldubai.ae/initiatives/ai-principles-ethics

[284] https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-uae#:~:text=Mainland%20UAE.or%20deployers%20of%20AI%20systems

ABNASIA.ORG

Hard Law

In the Mainland UAE, there is no single law regulating AI. However, several decrees have been passed on discrete issues:

1. In 2018, the Federal Decree Law No. 25 on the Project of Future Nature was issued – it allows the Cabinet to issue interim licenses for innovative projects being rolled out in the UAE that use AI, where there is no existing regulatory framework.[285]

2. In 2024, Law No. (3) of 2024 Establishing the Artificial Intelligence and Advanced Technology Council (AIATC) was issued, establishing a Council to regulate projects, investments and research related to artificial intelligence and advanced technology in the emirate of Abu Dhabi.

3. Existing sectoral laws may apply to AI – for example, the UAE Penal Code and UAE Federal Decree-Law No. 34 of 2021 on Combatting Rumors and Cybercrimes, as amended ('UAE Cybercrimes Law') might be applied to criminalize deepfakes, voice theft or IP infringement by AI.[286]

In the Financial Free Zones, no horizontal laws have been issued regulating AI. However, amendments have been made to existing data protection legislation that applies in the Dubai International Financial Centre. Article 10 of the Data Protection Regulations imposes certain obligations on deployers and operators of 'autonomous and semi-autonomous systems'.

Estonia

Soft Law / Regulatory Sandbox

Estonia's new AI Strategy (2022-2023) is a continuation of its first national AI strategy for 2019-2021. It aims to support 'regulat[ig] the development and use of AI in a human-centered and trustworthy way, i.e. in a reliable, ethical, and lawful way that respects fundamental rights, as well as to establish a set of rules on civil liability related to AI.'[287] The AI strategy was developed by a cross-sectoral taskforce including representatives from state authorities, the private sector, universities, and sectoral experts.[288]

The strategy mentions several specific actions to implement human-centric, trustworthy AI. For example, it requests the Ministry of Economic Affairs and Communications (MEAC), the Ministry of Justice (MK) and the Data Protection Inspectorate (DPI) to develop 'requirements and measures to support the development and use of human-centered and reliable AI solutions', and develop relevant policies to increase public trust and mitigate AI risks. Another suggestion is for Estonia to develop a fundamental rights impact assessment model and guidance materials.

In 2018, the Estonian minister for digital development signed a declaration on 'AI in the Nordic-Baltic region', establishing a collective framework for 'AI in the NordicBaltic region' establishing a collaborative framework on 'developing ethical and transparent guidelines, standards, principles and values to guide when and how AI applications should be used' and 'on the objective that infrastructure, hardware, software and data, all of which are central to the use of AI, are based on standards, enabling interoperability, privacy, security, trust, good usability, and portability.'[289]

Estonia has endorsed the OECD AI Principles and the UNESCO Recommendation on the Ethics of AI.

Hard Law

Estonia is a member of the EU and therefore the EU AI Act is applicable within its territory (for more on the EU AI Act, see box [x] above). Estonia also contributed to negotiations for the Council of Europe Framework Convention on AI, Human Rights, Democracy and the Rule of Law (discussed at section [x] above). As an EU member state, the Digital Services

[285] https://uaelegislation.gov.ae/en/legislations/1980

[286] https://insightplus.bakermckenzie.com/bm/data-technology/united-arab-emirates-deepfakes-and-the-use-of-artificial-intelligence-ai-legal-issues-and-considerations

[287] https://en.kratid.ee/_files/ugd/980182_e319a94450384ca198f027ba84/cbace.pdf

[288] https://f98cc689-5814-47ec-86b3-db505a7c3978.filesusr.com/ugd/7df26f_486454c9f32340b2820 6e140350159cf.pdf

[289] https://www.norden.org/en/declaration/ai-nordic-baltic-region

ABNASIA.ORG

Act also imposes some obligations on online intermediaries and platforms that use AI – for example, it prohibits targeted advertising based on a person's sexual orientation, religion, ethnicity, or political beliefs.

Other relevant laws applicable to AI systems include:

1. GDPR
2. EU Charter of Fundamental Rights
3. Council of Europe Convention 108+ for the protection of individuals with regard to the processing of personal data

Estonia is also considering a package of modifications to existing legal frameworks, separate to the EU AI Act.[290] These are intended to be targeted to solving specific problems that can be regulated independently of EU action.

[290] AIDV 2023 Report, p.443.

ABNASIA.ORG

WORLD BANK GROUP

ABNASIA.ORG