ITU Journal
Future and evolving technologies

ARTICLE

# Strengthening AI governance: International policy frameworks, security challenges, and ethical AI deployment

***Satyam Kumar*[1], *Priyansha Upadhyay*[1], *Gobi Ramsamy*[1]**

[1] *Christ University, Bangalore, India*

Corresponding author: Gobi Ramsamy, gobi.r@christuniversity.in

Generative AI is changing the very essence of digital interaction, affecting organizations and governments reaching the critical juncture where privacy vulnerabilities and security threats now evolve more swiftly than protective measures can be implemented, creating an urgent need to implement an overarching safeguard that can balance innovation with fundamental rights to sovereignty over data. This research examines the challenges through the lens of global preparedness, building on the insights from the analysis of the Global Index for Responsible AI (GIRAI) dataset. The analysis reveals a digital divide in how countries address privacy and security concerns in generative AI systems. While some nations have implemented robust frameworks to protect data privacy and ensure AI security, others lack essential safeguards, leaving their citizens vulnerable. The study investigates three critical dimensions: regulatory frameworks for responsible AI systems, technical safeguards against privacy breaches and security threats, and the collaborative role of governments, academia, industry, and civil society in establishing responsible practices. The findings emphasize the urgent need for international cooperation to establish a thorough, culturally-sensitive framework so that every human being, no matter where they are located, can benefit from responsible AI while being protected from the inherent privacy and security risks that generative AI poses.

Keywords: AI governance, AI safety and compliance, data privacy, ethical AI deployment, fully homomorphic encryption (FHE), generative AI (GAI), large language models (LLMs), responsible AI

## 1. INTRODUCTION

The emergence of highly sophisticated generative AI systems marks a paradigm shift in the digital environment: concomitant with novel articulations of power come unique challenges in privacy and security that existing governance frameworks struggle to address. Unlike previous AI technologies, generative models can process, synthesize, and reproduce vast quantities of information, including potentially sensitive data, while generating content that may circumvent traditional security protocols. This technological leap demands a re-imagining of how we protect individual privacy rights and organizational security in an era where AI can mimic human communication with unprecedented fidelity.

The privacy implications of generative AI extend far beyond traditional data protection concerns. These systems risk exposing confidential information through mechanisms such as memorization of training data, inference attacks that extract sensitive details, and the generation of personally identifiable information without proper consent or controls. Simultaneously, security vulnerabilities include sophisticated prompt manipulation, model poisoning through contaminated training data, and the automated creation of convincing disinformation or malicious code that can compromise systems at scale.

The concept of responsible AI governance, comprising ethical principles, technical standards, and regulatory frameworks, has become critical in the pursuit of these

technologies in healthcare, finance, education, and public services, to name a few [1]. However, the disparate implementation of such governance creates dangerous gaps in collective mitigation against risks from generative AI.

This research builds on the comprehensive Global Index for Responsible AI (GIRAI) 2024 edition dataset to analyze how different nations are responding to generative AI's privacy and security challenges [2]. An analysis of 138 countries concerning the selected issues yields unique insights into global preparedness, gaps in protection that need to be addressed, and the role of economic development in facilitating or hindering the implementation of effective safeguards for generative AI technologies. The analysis focuses particularly on three interconnected dimensions: the adaptation of existing regulatory frameworks to address generative AI's unique characteristics, the practical implementation of technical countermeasures against emerging threats, and the development of multi-stakeholder approaches that engage government, industry, academia, and civil society in creating responsible governance practices. The research intends through this exhaustive examination to contribute its results to the urgent discussion worldwide on how to utilize the transformative capabilities of generative AIs while counterpoising these interests against fundamental rights and security interests in various contexts around the globe.

### 1.1 Literature review

Recent research highlights the growing complexity of Artificial Intelligence (AI) regulation and its critical role in managing trust and risk across both private and public sectors. Alalawi et al. (2024) use evolutionary game theory to model the dynamics between users, AI developers, and regulators, showing that trust in AI is significantly strengthened when regulators are incentivized and users make informed decisions based on regulatory performance particularly when regulatory costs are kept manageable [3].

This aligns with broader concerns around AI governance, where managing not only immediate but also long-term and existential risks is paramount. Turchin and Denkenberger (2021) classify dozens of AI-related global catastrophic risks, from misaligned objectives to superintelligence scenarios, underscoring the absence of a one-size-fits-all solution to AI safety and the urgent need for proactive governance strategies [4].

In parallel, Zuiderwijk, Chen, and Salem's systematic literature review of AI in public governance identifies key challenges in legal, ethical, and social domains, advocating for sustained attention to transparency, accountability, and inclusive policymaking to support responsible AI use in government contexts [5]. Together, these studies emphasize the necessity of adaptive, incentive-aligned, and participatory frameworks that can address diverse AI risks while reinforcing public trust.

## 2. METHODOLOGY

### 2.1 Dataset description and coverage

The Global Index for Responsible AI (GIRAI) provides the first comprehensive framework for measuring and comparing AI governance readiness across 138 countries and jurisdictions. The research leverages GIRAI data to offer insights into the global state of AI governance with the following primary objectives:

- Assess the current state of AI governance across various regions and development levels.
- Identify critical gaps and disparities in AI governance frameworks.
- Understand the relationship between economic development and AI governance readiness.
- Provide evidence-based recommendations for enhancing global AI governance.

**Table 1** – Thematic areas across dimensions

| Dimension | Thematic Area |
|---|---|
| Responsible AI Governance | National AI Policy<br>Impact Assessments<br>Human Oversight and Determination<br>Responsibility and Accountability<br>Proportionality and Do No Harm<br>Public Procurement<br>Transparency and Explainability<br>Access to Remedy and Redress<br>Safety, Accuracy, and Reliability |
| Human Rights & AI | Gender Equality<br>Data Protection and Privacy<br>Bias and Unfair Discrimination<br>Children's Rights<br>Labor Protection and Right to Work<br>Cultural and Linguistic Diversity |
| Responsible AI Capacities | Competition Authorities<br>Public Sector Skills Development<br>International Cooperation |

The GIRAI framework is built on three fundamental pillars that offer a holistic view of a country's AI governance capability:

- **Government frameworks (40%):** Evaluates the legislative and regulatory foundation for AI governance.
- **Government actions (40%):** Assesses the practical implementation of AI policies and initiatives.

- **Non-state actors (20%):** Measures the engagement of academic institutions, the private sector, and civil society in AI governance.

These pillars are further subdivided into multiple dimensions: responsible AI governance, human rights & AI, and responsible AI capacities, which are decomposed into 18 thematic areas and 57 indicators, as stated in Table 1. This structure addresses key issues such as data privacy, children's rights, labor protection, and AI system transparency, enabling a detailed yet broad assessment of national AI ecosystems.

## 2.2 Data collection methods

Data for the study was gathered through a coordinated Global Research Network consisting of 11 Regional Research Hubs (RRHs) located at prominent AI and technology research organizations worldwide [2]. The network has recruited a total of 138 national level researchers, where each RRH appoints a team leader for local data collection and continuous training and review for quality control. The data collection elements include the following:

- **Recruitment and coordination:** Every RRH managed regional recruitment and training of country-level researchers, providing training and a complete researcher's handbook on the conceptual components, dimensions, thematic areas, and indicators of the Global Index on Responsible AI.
- **Training and capacity building:** The core team of GIRAI conducted several extensive training workshops that imparted knowledge on the use of survey instruments, quality assurance protocols, and review procedures thereby ensuring all the survey data-gatherers would consistently and reliably do their work.
- **Primary data survey tool:** Data was collected using an online survey platform (Survey Solutions, developed by the World Bank), where researchers completed structured questionnaires featuring standardized, closed-ended questions. Each response was supported by evidence and subsequently reviewed by team leaders, a board of external reviewers, and a core research team.

To further elaborate on the nature of the questionnaire, these structured, closed-ended questions were not administered in a vacuum. Each thematic area and its constituent indicators were accompanied by comprehensive contextual information and detailed guidance for the researchers, as outlined in the researcher's handbook and reinforced during training. For instance, within a thematic area such as 'Gender Equality' under the 'Human Rights & AI' dimension, researchers were provided with precise definitions of core concepts (ex: 'Frameworks' referring to governmental guidelines, policies, or regulations), explanations of their significance, illustrative examples of qualifying evidence, and specific instructions on how to locate such evidence. This methodological depth ensured that while the questions themselves were standardized for comparability, the responses were deeply informed and consistently interpreted across the diverse national contexts, allowing for a robust assessment of whether governmental policies, regulations, or guidelines adequately addressed the specific criteria of each indicator.

This structured approach ensured robust, reliable, and comparable data across all participating countries.

## 2.3 Analysis framework

The framework of analysis underwent several critical stages, thus facilitating a sound evaluation of responsible AI governance:

- **Data cleaning and preprocessing:** Data was brought in from Excel files, while other non-numeric entries were coerced to missing values, which were substituted later by zeros or with an appropriate mean value of the corresponding geographical subregion.
- **Country classification:** Countries were categorized into developed, developing, and underdeveloped groups based on a predefined classification system. This facilitated tailored analysis of AI governance maturity across diverse economic contexts.
- **Indicator calculation and normalization:** The Global Index is structured into 3 dimensions, 3 pillars (government frameworks, government actions, and non-state actors), 19 thematic areas, and 57 indicators. Qualitative survey responses were quantitatively scored using pre-established rules. Each indicator was then normalized to a common 0–100 scale using a minimum-maximum formula, ensuring consistency across different indicator types.
- **Adjustment using country-specific coefficients:** Normalized indicator values were adjusted by country-specific coefficients derived from metrics such as regulatory quality, rule of law, government effectiveness, control of corruption, and freedom indices. These coefficients were rescaled with a minimum bound of 0.25 to ensure that each country retains at least 25% of its original score, even when contextual metrics are lower.
- **Score aggregation:**
  - **Pillar scores:** For each country, the scores for the three pillars were calculated as the arithmetic average of the 19 thematic area indicator scores.
  - **Index score:** The overall Global Index score was computed as a weighted average of the three pillar scores, with weights assigned as follows: 40% for government frameworks, 40% for government

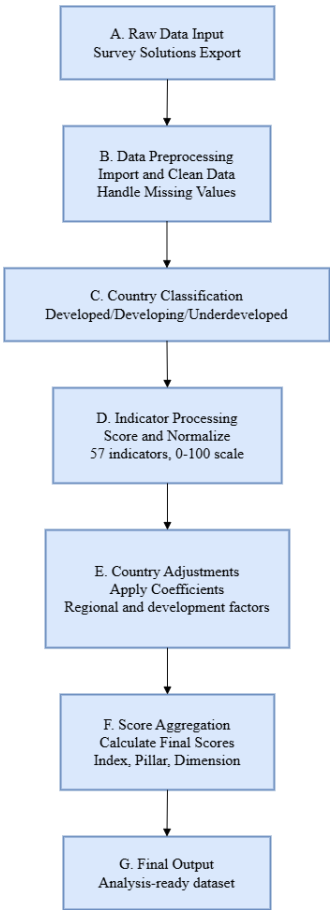



**Figure 1** – Analysis framework

actions, and 20% for non-state actors.

– **Thematic area and dimension scores:** Different aggregation procedures yielded scores across various thematic areas and dimensions, allowing for a detailed assessment of the aspects of responsible AI governance.

## 2.4 Visualization techniques

To elucidate the findings and facilitate comparative analysis, several visualization techniques were employed:

- Heatmaps to depict global AI preparedness and highlight performance across different thematic areas and development statuses.
- Box plots to illustrate the distribution of AI governance scores among developed, developing, and underdeveloped countries.
- Radar plots for detailed comparisons of key countries across multiple dimensions of AI governance.
- Bar charts to display regional average index scores, revealing disparities in governance maturity across global regions.

## 2.5 Limitations and assumptions

The methodological approach encompasses certain limitations and assumptions. Reducing complex national AI ecosystems into the framework of scores can help with comparative analysis, but it might also act as an oversimplification of intricate nuances and cross-cutting issues that bear on AI governance. In addition, publicly available data introduce variations in information regulation, accessibility to resources, and currency of data between countries, which, in turn, could bias the outcome toward government transparency instead of real efficacy of responsible AI practices. Additionally, the Global Index primarily measures the existence of AI governance frameworks and related actions without directly assessing their effectiveness in protecting and promoting human rights or evaluating the human rights impact of specific AI systems in various contexts. The use of peer group mean values to impute missing data may introduce biases, especially in regions with limited or less comprehensive data. Finally, the adopted weighting scheme (40% for frameworks and government actions, 20% for non-state actors) and the normalization process are based on theoretical assumptions that may not fully capture all relevant dimensions of responsible AI governance.

# 3. RESULTS

## 3.1 Global landscape analysis

As evidenced by data from the Global Index on Responsible AI, the global scene of responsible AI governance shows considerable disparities in preparedness to confront privacy and security challenges. As visualized in region-wise average score bar graph in Fig. 2, a prominent "governance canyon" separates nations with functioning modern AI governance from those that barely have minimum governance structures, a mirror closely reflecting existing socioeconomic inequalities. The advanced countries had been rated with an average of 64.16 while maintaining relatively elaborated frameworks, but these apparent leaders are within the challenges of realizing ideal governance standards. The developing ones, with an average score of 28.46, show frameworks emerging inconsistently, while the underdeveloped countries, with an average score of 2.86, hardly exhibit any governance structures at all, giving rise to myriad vulnerabilities. The disparities are worrisome given the expeditious advancement of generative AI equitably tending to risk aggravating such social and economic imbalances and announcing the dawn of yet another chapter of global inequity.

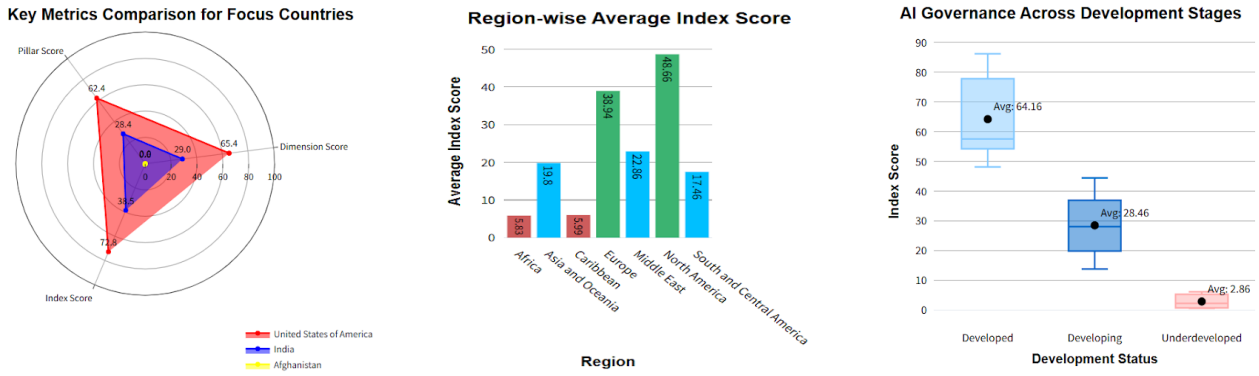



**Figure 2** – Analysis visualizations

### 3.1.1 Regional comparisons and variations

**Top-tier regions:**

- With an average index score 48.66, North America allows for strong data protection and transparency frameworks.
- Strong international cooperation (79.56) and data protection (74.79), Europe is slightly behind, with an average score of 38.94.

**Middle-tier regions:**

- The Middle East (22.86), Asia and Oceania (19.8), and South and Central America (17.46) exhibit moderate strengths in international cooperation but struggle with responsibility frameworks and accountability mechanisms.

**Bottom-tier regions:**

- Critical gaps in children's rights, protections and responsibility frameworks, Africa (5.83), and the Caribbean (5.99) face significant challenges. Thus, geographic proximity to high-scoring regions does not guarantee improved governance, as systemic issues persist.

### 3.1.2 Leading countries and emerging players

Advanced countries like the Netherlands (86.16), Germany (82.77), and the United States (72.8) have set benchmarks for comprehensive governance, which includes mature frameworks for privacy, data protection, and transparency. Emerging stars like Estonia (67.61), India (38.5), and Brazil (Rank 18) are fast developing their own AI governance structures. Countries use partnerships and knowledge transfer to develop their systems, primarily to balance innovation with security concerns.

### 3.1.3 Identification of gaps and challenges

The analysis identifies several critical gaps in global AI governance:

- **Governance gaps:**
  - *Children's rights and protections:* Underdeveloped nations score near zero, leaving young populations vulnerable.
  - *Data protection disparities:* Scores range from 74.79 in developed countries to just 6.79 in underdeveloped nations.
  - *Responsibility and accountability frameworks:* The absence of frameworks in underdeveloped nations poses severe risks to ethical AI deployment.
- **Systemic challenges:**
  - *Implementation gap:* Even in countries with strong policies, practical implementation still lags.
  - *Resource constraints:* Significant investments do not necessarily translate to effective governance, as seen in countries like Saudi Arabia (28.95).
  - *Global standards alignment:* Vast differences in governance capacities hinder the implementation of universal AI standards.

International cooperation emerges as a critical, consistently prioritized area across all development levels, with scores of 79.56, 52.21, and 17.56 for developed, developing, and underdeveloped nations, respectively. Without targeted interventions, these disparities risk creating a world where the privacy and security benefits of generative AI are accessible only to select nations, leaving others increasingly vulnerable.

## 3.2 Thematic analysis

The Global Index on responsible AI shows significant divergences in various thematic areas of privacy and security concerns surrounding Responsible AI, as shown in Fig. 3. The data indicates a clear line between countries with strong frameworks and those with severe vulnera-

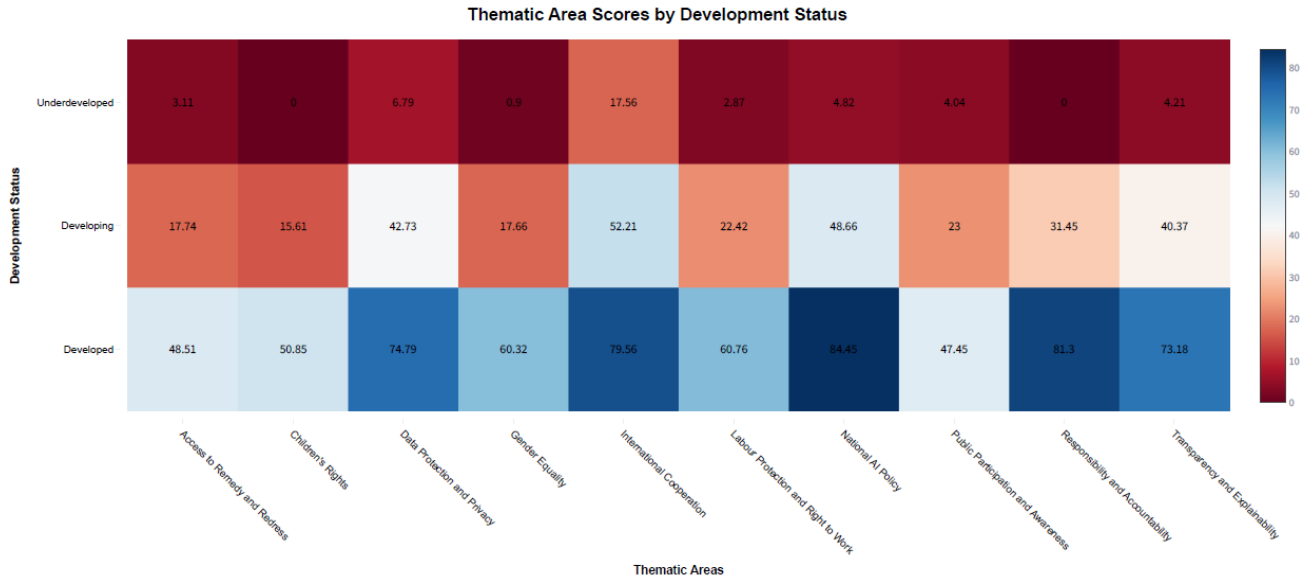



**Figure 3** – Thematic analysis

bilities to AI socio-economic disparities. Key thematic areas are evaluated as follows:

- **Data protection and privacy:** Developed nations score 74.79, reflecting comprehensive legislation (ex: GDPR) [6], while developing and underdeveloped regions score 38.45 and 6.79. The lower scores highlight varying levels of consent and protection for training data.
- **Transparency and explainability:** Advanced in developed regions (73.18), facilitating scrutiny of AI decision-making. However, this is inconsistent in developing countries (35.22) and nearly absent in underdeveloped regions (1.89), which undermines the ability to effectively audit AI systems for security vulnerabilities or biases.
- **Responsibility and accountability:** Established frameworks in developed nations (59.42) provide clearer pathways for addressing AI-related harms. This contrasts sharply with limited mechanisms in developing regions (31.45) and a complete absence in underdeveloped nations (0.00), creating significant governance gaps and unaddressed risks when AI systems cause negative outcomes.
- **International cooperation:** Analysis suggests a promising avenue for coordinated global action with encouragingly high scores across the board (79.56 in developed, 52.21 in developing, and 17.56 in underdeveloped regions).
- **Children's rights:** Although developed nations show moderate protection (45.67), underdeveloped regions score 0.00, indicating a severe lack of specific safeguards. This deficiency exposes vulnerable child populations to heightened risks of privacy violations and exposure to harmful algorithmically-generated content.

### 3.2.1 *Cross-theme correlations and global disparities*

The analysis reveals significant correlations across key governance themes:

- Data protection and transparency exhibit an almost perfect correlation (0.94), indicating that strong privacy frameworks often evolve alongside transparent processes.
- A strong correlation (0.87) between responsibility and technical robustness suggests that well-defined accountability mechanisms enhance AI security.
- A moderate correlation (0.71) between international cooperation and public participation highlights the importance of global collaboration in fostering diverse stakeholder engagement.

Success stories from leading AI governance models offer insights into best practices, the Netherlands and Germany set regulatory benchmarks through GDPR and the AI Act [7], Estonia excels in digital transformation via strong public-private partnerships, Singapore leads in technical robustness through certification and research, and Brazil showcases rapid progress through international collaborations. However, persistent governance gaps remain, particularly in underdeveloped regions where issues such as children's rights and responsibility receive minimal policy attention, and even advanced nations struggle with data protection and technical security implementation. To bridge these gaps, targeted interventions are necessary, including harmonized cross-border data governance, development of compliance tools, and stronger certification frameworks, ensuring equitable and responsible AI deployment worldwide.

**Table 2** – Global AI governance landscape: Key features and challenges

| Country/Region | Key AI Governance Features | Challenges |
|---|---|---|
| United States | Federal initiatives like the National AI Initiative Act, strong industry participation, and AI research funding. | Privacy laws remain fragmented, and bias mitigation challenges persist. |
| European Union | The AI Act ensures accountability while balancing innovation and privacy. Cross-border policy harmonization strengthens governance. | Implementation complexity and compliance burdens for businesses. |
| China | State-driven AI governance with strict control over ethics and applications. | Concerns over state surveillance and limited data privacy protections. |
| India | AI policies emphasize ethics, inclusivity, and economic growth. | Absence of a dedicated AI law and fragmented regulatory oversight. |
| Africa (Selected Nations) | Relies on global frameworks and external collaborations for AI governance. | Limited regulatory capacity and resource constraints hinder AI adoption and ethical compliance. |

## 3.3 Pillar-wise deep dive

A comprehensive analysis of responsible AI governance requires examining its foundational pillars, each of which contributes uniquely to a nation's AI maturity. These pillars include regulatory frameworks, government actions, and non-state actor participation. By assessing their individual strengths and interdependencies, we can better understand the disparities in AI governance across different regions.

### *3.3.1 Analysis of each foundational pillar*

**Regulatory frameworks:** Robust AI regulations are rooted in ethical principles, data privacy, and security codes. The regulations in most developed nations are well-defined, with independent oversight and enforceable AI ethics guidelines. On the contrary, there are comprehensive, coherent arrangements regarding legislation that lack effective enforcement mechanisms and have gaps, in many ways, in most developing regions.

**Government actions:** Governments play an essential role in AI development because of policy initiatives, funding, and cross-border collaboration. A country exhibiting a high level of AI governance maturity would invest in a national AI strategy, public-private partnerships, and programs for developing talent. Limited technical knowledge and infrastructure, however, hinder the progress of developing countries.

**Non-state actor participation:** The roles of academia, industry, and civil society organizations in shaping AI governance are significant. In developed countries, think tanks, advocacy groups, and AI research centers are active players in influencing policy decisions. In contrast, in underdeveloped areas, institutional capacity may be lacking, thereby limiting the involvement of stakeholders in AI policymaking.

### *3.3.2 Inter-pillar relationships*

These three pillars are highly interlinked and reinforce each other. Strong regulatory frameworks make state intervention more effective, while actors outside the state influence transparency and accountability. Weaknesses in governance structures impose challenges where one area of the three pillars falters, e.g.: ineffective regulation and implementation of government-led AI programs and industry participation for responsible concerns are hampered. For example, the European Union's AI Act benefits from strong policy backing, governmental oversight, and active industry compliance, making it a global benchmark for AI regulation. Conversely, in regions where legal frameworks are underdeveloped, AI initiatives remain fragmented, and ethical AI adoption faces significant challenges.

AI governance is shaped by regulatory frameworks, government actions, and non-state actor participation, each with distinct drivers and challenges. Though regulatory frameworks such as GDPR, HIPAA [8], and international AI standards advocate compliance, various very slow policy adoption, and weak enforcement hinder achieving it. Due to budget constraints and bureaucratic inefficiencies, government activities are not much facilitative, even regarding allocations for AI investments and cross-border collaboration. Non-state actors like research institutions and corporate ethics programs advocate responsible AI but often lack transparency and access to governance processes. Strengthening these pillars is crucial for secure and equitable AI governance.

#### 3.3.3 Country-specific case studies

As can be understood from Table 2, the top-tier regions, namely North America and Europe, demonstrate a clear correlation between robust AI governance frameworks and leadership in Large Language Model [9] development. North America, with an average governance index score of 48.66, has produced industry-leading models including OpenAI's GPT-4 [10], Anthropic's Claude [11], Google's Gemini [12], and Meta's Llama [13], while Europe (38.94) has fostered distinctive approaches through models like Aleph Alpha's Luminous [14] in Germany (82.77), Mistral AI [15] in France. These regions leverage their governance maturity to create AI ecosystems that balance innovation with responsible oversight, enabling both commercial success and technical excellence.

By contrast, regions with governance scores below 20, such as Africa (5.83) and the Caribbean (5.99), face a "governance canyon" that manifests as technological dependency. Middle-tier nations have found success through specialized focus, India (38.5) has developed targeted solutions like AI4Bharat's Bhashini for Indian languages [16] and Sarvam AI's OpenHathi [17], while Singapore's high technical standards have enabled AI Verify Foundation's testing toolkit and specialized medical models like NUS's MedGPT [18]. China, despite its unique governance approach, has produced a competitive model such as DeepSeek-R1 [19]. This pattern suggests that strengthening AI governance frameworks is not just an ethical consideration but a strategic necessity for developing indigenous LLM capabilities that reflect local languages, cultures, and priorities.

## 4. DISCUSSIONS

### 4.1 Policy implications and international cooperation

The study suggests an urgent need for comprehensive policy interventions to address the varying disparities in responsible AI governance. Policymakers must develop robust regulatory frameworks, promote ethical AI deployment, and enhance international collaboration to ensure equitable governance across regions. Therefore, a multi-stakeholder approach, which accounts for the governments, business leaders, academia, and civil society, is necessary in promoting responsible and secure implementation of AI.

#### 4.1.1 Key recommendations for policymakers

Clear AI governance frameworks are ultimately critical for realizing global standards and ensuring secure AI deployment, moving beyond general compliance to address specific AI risks. Policymakers must develop robust regulatory frameworks that mandate specific security considerations throughout the AI lifecycle. For example, the European Union's AI Act aims to establish a human-centric, resilient, and trustworthy AI ecosystem, where resilience against attacks and unintended behavior is a core tenet for high-risk systems [20]. This approach is echoed in various national strategies that increasingly adopt comprehensive AI risk management frameworks, such as the one developed by NIST in the United States, which guides organizations in managing AI risks, including those related to security and robustness [21]. These frameworks emphasize the need for AI systems to be secure by design.

Institutionally, this means strengthening oversight and enforcement. This includes establishing or empowering bodies focused on AI safety and security, such as the UK's AI Safety Institute, which conducts research and evaluations of advanced AI models. Such institutions can drive the adoption of methodologies for rigorous testing, red-teaming, and continuous monitoring of AI systems to uncover and mitigate vulnerabilities before and after deployment, addressing the practical attack and defense landscapes of modern AI [22]. Furthermore, the development and promotion of AI auditing practices are essential. Model-based auditing, for instance, can provide systematic ways to assess AI systems against security requirements and ethical guidelines, offering a pathway to verifiable compliance, a concept gaining traction globally [23].

Public-Private Partnerships (PPPs) should be actively encouraged to drive responsible and secure AI innovation. This includes incentivizing the integration of robust security safeguards and Privacy-preserving Technologies (PETs) like Fully Homomorphic Encryption (FHE) and differential privacy from the AI system's design phase [24] [25]. Many advanced research programs, often government-backed, are exploring how PETs can be practically implemented to protect sensitive data used in AI training and inference, thereby minimizing security risks associated with data breaches or misuse. Countries leading in AI are increasingly recognizing that fostering an ecosystem where PETs are both developed and deployed is crucial for building trust and ensuring data security in AI applications. By learning from these state-of-the-art implementations, such as comprehensive risk management, dedicated safety institutions, advanced model security practices, PET integration, and robust auditing mechanisms, policymakers can create effective, adaptive, and legally-enforceable AI governance.

#### 4.1.2 *Areas requiring international cooperation*

To underpin global AI governance and regulation, some areas will necessitate international cooperation. Global AI standards and interoperability need to be prioritized as institutions such as OECD [26], IEEE [27], and ISO [28] make groundwork efforts to set up standard AI governance protocols that cater to cross-border deployments. Data privacy and security frameworks should not be limited to regional regulations such as GDPR and HIPAA, and this will require the drafting of global treaties related to AI data-sharing, encryption, and cybersecurity. Collaboration with privacy-enhancing AI techniques such as FHE and secure multi-party computation [29] should be supported in a way that would provide for increased AI security. Finally, these projects, led by the United Nations and human rights organizations, should strengthen ethical AI and human rights protection by establishing AI policies that protect against discrimination, surveillance abuse, and algorithmic bias and promote fundamental human rights principles.

### 4.2 Future outlook

Governance frameworks need to be flexible to address emerging challenges and technological progress. It is essential to implement risk-based AI regulation, where countries classify AI applications based on their impact on society. AI systems that pose a risk would be some that are involved in the areas of healthcare, finance, or law enforcement, and therefore must be subject to oversight and auditing in order to be deployed ethically. The bolstering of transparency and explainability requirements will strengthen public trust and assist regulatory compliance, thus forcing AI developers to deliver explainability reports for their models. Moreover, ensuring ethical AI should entail advancing the Fairness, Accountability, Transparency, and Ethics (FATE) principles through bias audits/policies, diverse datasets, and inclusive policies.

Future policy making would have to look at the big ticket items such as Generative AI and LLM regulation, deepfakes and misinformation, and biases in automated decision-making, among other things. Labor markets would be upended by AI automation, which means that investments in reskilling programs need to prepare future workers for an AI economy. Sustainability, of course, should be at the forefront, with policies mandating energy-efficient AI models and green AI initiatives that could ease the effects on the environment. Besides, in the area of national security, future policies will have to include a fine balance between innovation in AI and threat to the security that that creates, norms for responsible regulation of AI in cybersecurity, defense applications, and autonomous systems of usage being established. These best governance practices and future orientation policies will go far in establishing a transparent, equitable, and safe AI governance framework within which technological innovation can flourish but ethical and responsible deployment of AI be secured worldwide.

## 5. CONCLUSIONS

The study shows considerable global differences in responsible AI governance, which directly impacts privacy and security concerning generative AI systems. A main takeaway is that AI governance is not a matter of one form fitting all. Policies should be context-specific, flexible enough to factor in the dynamics of a certain technological and regulatory landscape in which a country operates. The study demonstrates the interlinked nature of the various pillars of AI governance, wherein loopholes in any particular sector, such as policy enforcement, can hamper efforts towards AI accountability and securing systems. While the developed countries have set up strong frameworks with legal scrutiny, from the perspective of compliance by the industry, the underdeveloped regions continue to face utter disaster in terms of governance, especially in the areas of data protection and children's rights. Being an extension of existing socio-economic inequalities, such a form of the difference can be characterized as a "governance canyon" from the point of view of humanity. Studies demonstrate that, indeed, embodiments of AI governance should be context-driven by each of the country regulations, with international collaboration being a promising strategy to close these governance gaps.

As AI capabilities develop, the governance gaps will deepen social and economic disparities. What we are left with is no mere account of these inequalities but rather an urgent call to action. The data suggests that we are not measuring numbers, we are witnessing the early beginnings of a new form of global inequality that could, if left unabated, characterize the next era of human development.

## ACKNOWLEDGEMENT

The authors would like to thank all the contributors and researchers involved in the Global Index for Responsible AI project.

## REFERENCES

[1] J. M. Helm, A. M. Swiergosz, H. S. Haeberle, et al. "Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions". In: *Current Reviews in Musculoskeletal Medicine* 13 (2020), pp. 69–76. DOI: 10.1007/s12178-020-09600-8. URL: https://doi.org/10.1007/s12178-020-09600-8.

[2] Global AI Governance Index. *Global AI Governance Index*. Accessed: 2025-03-02. 2025. URL: https://www.global-index.ai/.

[3] Zainab Alalawi, Paolo Bova, Theodor Cimpeanu, Alessandro Di Stefano, Manh Hong Duong, Elias Fernandez Domingos, The Anh Han, Marcus Krellner, Bianca Ogbo, Simon T. Powers, and Filippo Zimmaro. *Trust AI Regulation? Discerning users are vital to build trust and effective AI regulation*. 2024. arXiv: 2403.09510 [cs.AI]. URL: https://arxiv.org/abs/2403.09510.

[4] A. Turchin and D. Denkenberger. "Classification of global catastrophic risks connected with artificial intelligence". In: *AI Soc* 35 (Mar. 2020), pp. 147–163. DOI: 10.1007/s00146-018-0845-5.

[5] Anneke Zuiderwijk, Yu-Che Chen, and Fadi Salem. "Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda". In: *Government Information Quarterly* 38.3 (2021), p. 101577. ISSN: 0740-624X. DOI: https://doi.org/10.1016/j.giq.2021.101577. URL: https://www.sciencedirect.com/science/article/pii/S0740624X21000137.

[6] GDPR-Info.eu. *General Data Protection Regulation (GDPR)*. URL: https://gdpr-info.eu/.

[7] European Commission. *The AI Act*. 2021. URL: https://artificialintelligenceact.eu/wp-content/uploads/2021/08/The-AI-Act.pdf.

[8] Centers for Disease Control and Prevention (CDC). *Health Insurance Portability and Accountability Act of 1996 (HIPAA)*. URL: https://www.cdc.gov/phlp/php/resources/health-insurance-portability-and-accountability-act-of-1996-hipaa.html.

[9] Shervin Minaee, Tomas Mikolov, Narjes Nikzad, Meysam Chenaghlu, Richard Socher, Xavier Amatriain, and Jianfeng Gao. *Large Lan-guage Models: A Survey*. 2024. arXiv: 2402.06196 [cs.CL]. URL: https://arxiv.org/abs/2402.06196.

[10] OpenAI et al. *GPT-4 Technical Report*. 2024. arXiv: 2303.08774 [cs.CL]. URL: https://arxiv.org/abs/2303.08774.

[11] Anthropic. *Claude 3 Model Card*. Accessed: 2025-03-02. 2024. URL: https://www-cdn.anthropic.com/de8ba9b01c9ab7cbabf5c33b80b7bbc618857627/Model_Card_Claude_3.pdf.

[12] Gemini Team et al. *Gemini: A Family of Highly Capable Multimodal Models*. 2023. arXiv: 2312.11805 [cs.CL].

[13] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. *LLaMA: Open and Efficient Foundation Language Models*. 2023. arXiv: 2302 .13971 [cs.CL]. URL: https://arxiv.org/abs/2302.13971.

[14] Aleph Alpha. *Luminous Explore: A Model for World-Class Semantic Representation*. Accessed: 2025-03-02. 2025. URL: https://aleph-alpha.com/luminous-explore-a-model-for-world-class-semantic-representation/.

[15] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. *Mistral 7B*. 2023. arXiv: 2310.06825 [cs.CL]. URL: https://arxiv.org/abs/2310.06825.

[16] Sumanth Doddapaneni, Rahul Aralikatte, Gowtham Ramesh, Shreya Goyal, Mitesh M. Khapra, Anoop Kunchukuttan, and Pratyush Kumar. *Towards Leaving No Indic Language Behind: Building Monolingual Corpora, Benchmark and Models for Indic Languages*. 2023. arXiv: 2212.05409 [cs.CL]. URL: https://arxiv.org/abs/2212.05409.

[17] Jay Gala, Thanmay Jayakumar, Jaavid Aktar Husain, Aswanth Kumar M, Mohammed Safi Ur Rahman Khan, Diptesh Kanojia, Ratish Puduppully, Mitesh M. Khapra, Raj Dabre, Rudra Murthy, and Anoop Kunchukuttan. *Airavata: Introducing Hindi Instruction-tuned LLM*. 2024. arXiv: 2401.15006 [cs.CL]. URL: https://arxiv.org/abs/2401.15006.

[18] Zeljko Kraljevic, Anthony Shek, Daniel Bean, Rebecca Bendayan, James Teo, and Richard Dobson. *MedGPT: Medical Concept Prediction from Clinical Narratives*. 2021. arXiv: 2107.03134 [cs.CL]. URL: https://arxiv.org/abs/2107.03134.

[19] DeepSeek-AI et al. *DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning*. 2025. arXiv: 2501 .12948 [cs.CL]. URL: https://arxiv.org/abs/2501.12948.

[20] L. Floridi. "The European Legislation on AI: a Brief Analysis of its Philosophical Approach". In: *Philos. Technol.* 34 (June 2021), pp. 215–222. DOI: 10.1007/s13347-021-00460-9.

[21] National Institute of Standards and Technology (NIST). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. Tech. rep. NIST AI 600-1. NIST Trustworthy and Responsible AI. National Institute of Standards and Technology, U.S. Department of Commerce, July 2024. DOI: 10.6028/NIST.AI.600-1. URL: https://doi.org/10.6028/NIST.AI.600-1.

[22] Marcus Comiter. *Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It*. Tech. rep. Belfer Center for Science and International Affairs, Harvard Kennedy School, Aug. 2019. URL: https://www.belfercenter.org/publication/AttackingAI.

[23] J. Mökander, J. Schuett, H.R. Kirk, and et al. "Auditing large language models: a three-layered approach". In: *AI Ethics* 4 (Nov. 2024), pp. 1085–1115. DOI: 10.1007/s43681-023-00289-2.

[24] Alexander Viand, Christian Knabenhans, and Anwar Hithnawi. *Verifiable Fully Homomorphic Encryption*. 2023. arXiv: 2301.07041 [cs.CR]. URL: https://arxiv.org/abs/2301.07041.

[25] P. Jain, M. Gyanchandani, and N. Khare. "Differential privacy: its technological prescriptive using big data". In: *Journal of Big Data* 5 (2018), p. 15. DOI: 10.1186/s40537-018-0124-9. URL: https://doi.org/10.1186/s40537-018-0124-9.

[26] Organisation for Economic Co-operation and Development (OECD). *OECD Official Website*. URL: https://www.oecd.org/.

[27] Institute of Electrical and Electronics Engineers (IEEE). *IEEE Official Website*. URL: https://www.ieee.org/.

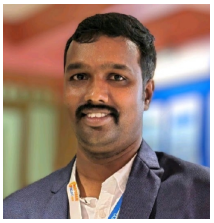
[28] International Organization for Standardization (ISO). *ISO Official Website*. URL: https://www.iso.org/.

[29] Brian Knott, Shobha Venkataraman, Awni Hannun, Shubho Sengupta, Mark Ibrahim, and Laurens van der Maaten. *CrypTen: Secure Multi-Party Computation Meets Machine Learning*. 2022. arXiv: 2109.00984 [cs.LG]. URL: https://arxiv.org/abs/2109.00984.

## AUTHORS

Satyam Kumar is a data enthusiast, working on projects related to LLMs and deep learning, with a strong focus on research, innovation, and practical implementations, aiming to push the boundaries of AI.

Priyansha Upadhyay is a data science specialist with expertise in artificial intelligence and machine learning. Holding a Master's degree in artificial intelligence, she excels at solving complex data challenges and implementing advanced solutions across multiple domains.



Gobi Ramsamy is an associate professor, committed to advancing computer science education and fostering interdisciplinary research. His work aligns with Sustainable Development Goals (SDGs) 04 (Quality Education) and 10 (Reduced Inequalities), focusing on inclusive and equitable technological advancements. He teaches Android app development, data analytics, and visualization to master's students at Christ University, Bangalore. As a responsible AI advocate, he actively promotes ethical AI development and application. He is also a Best Paper Award winner at ITU Kaleidoscope 2024.